November, 2009

Evaluating Economic Development Programs using Matched Employee-Employer data in a Quasi- Experimental Framework

Henry C Renski, *University of Massachusetts - Amherst*

Evaluating Economic Development Programs using Matched Employee-Employer data in a Quasi-Experimental Framework

Henry Renski, PhD
University of Massachusetts Amherst
Department of Landscape Architecture and Regional Planning

Working paper
Version date: October, 2009

Abstract

In the wake of shrinking public coffers, policy makers are demanding greater accountability from their economic development initiatives.   In a discipline known for 'claiming anything that falls,' attempts to objectively evaluate economic development programs have been stymied by ill-suited data sources and methods. Survey research is expensive and responding firms have an incentive to lie about the effectiveness of subsidies. Publicly available data on employment, wages, and other outcomes are highly aggregated and lack the power to capture impacts from anything other than the most dramatic, large-scale initiatives. Confidential employee- and establishment-level (micro) data holds considerable promise for more rigorous and objective approaches to program evaluation, but have rarely been used for this purpose.

This paper uses data on employee and employer characteristics to evaluate the short-term impacts of a training-subsidy initiative in the State of Maine. I match employee records in the state's Unemployment Insurance database to establishment records in the Quarterly Census of Employment and Wages to create a longitudinal database containing both firm and worker characteristics. This database is matched to program information on participating businesses and trainees and analyzed using a quasi-experimental approach with propensity-score matching. I found some evidence of a small positive impact on participant earnings and employment levels one year after training, but no significant immediate impacts.

Background

In the early months of 2007, the US Department of Labor's Employment and Training Administration (ETA) designated Coastal Maine as one of seven regional partnerships under its pilot Workforce Innovations and Regional Economic Development (WIRED) initiative. Soon thereafter, Maine formed the North Star Alliance Initiative (NSAI) to administer the grant with the goal of strengthening the competitiveness of the state's Marine Trades cluster.

Under the financial and technical support of WIRED, the NSAI worked with local educational institutions and industry representatives to establish industry-targeted training programs. Unlike most existing workforce development initiatives, which are designed to help dislocated and targeted populations secure paid employment, the training programs of the NSAI are largely targeted to employers and the employed labor force. Crossing the traditional division between economic and workforce development initiatives, the programs of the NSAI were designed with the goal of increasing business competitiveness and innovation (and thus contributing to long-term job creation) by expanding training opportunities in the use of cutting-edge production technologies and emerging product markets. To incentive participation, the NSAI offered employer-based subsidies to offset tuition and other training expenses. The NSAI sponsored a number of workforce development programs—ranging from financial support and training opportunities for instructors and vocational institutions, to short-courses and seminars in business planning and emerging market opportunities targeted to business managers, to educational scholarships for students pursing degrees in fields critical to the long-term needs of the industry. However, the largest of the NSAI workforce development programs consisted of intensive short-term (one to two weeks) certificate programs primarily targeted to incumbent employees. The NSAI also provided training subsidies for recent hires through its on-the-job-training (OJT) program. This study focuses solely on the outcomes from OJT and incumbent certificate programs.

Developing effective and innovative tools for monitoring and evaluating program outcomes was a secondary aspect of the NSAI agenda. As part of its program evaluation strategy, the NSAI built a comprehensive database of all Marine Trades companies in the state. It then partnered with the Maine Department of Labor's Center for Workforce Research and Information (CWRI) to monitor changes in employment and wages among NSAI-supported businesses and trainees using CWIR's confidential employer and employee databases that are collected as part of the state's unemployment insurance program.

The purpose of this report is to demonstrate the use of confidential unemployment insurance data for evaluating economic and workforce development programs. While every state collects this data, it is rarely used for program evaluation and monitoring. This is, in part, because of strict disclosure rules that prohibit reporting information that might possibly reveal the characteristics of individual businesses or workers. These databases also cannot be used to determine whether specific companies and workers are in compliance of government program rules or regulations. For example, they cannot be used to determine whether a particular company has met wage or employment quotas often associated with publicly subsidies

incentives. However, statistical uses of these confidential databases are allowed under BLS rules, making it possible to evaluating the overall effectiveness of program activities.[1]

This report opens with a brief discussion of the use of quasi-experimental approaches in economic development policy evaluation. In the next section, I describe the process for building an employer-employee longitudinal file by merging program data with confidential databases of establishment and worker characteristics. The next two sections evaluate the short-term impacts of NSAI training by comparing changes in employment and wages among businesses and workers who participated in NSAI training programs to similar businesses and workers that did not. I close with a brief summary of key findings, a discussion of the limitations of this approach, and future directions for this study.

Measuring program impacts using quasi-experimental design methods
One of the greatest challenges in evaluating public policy is addressing the counterfactual scenario of what would have happened in the program had not been initiated (Bartik 2002). In a laboratory setting, this counterfactual is satisfied through the random assignment of treatments, which creates a probabilistic equivalence between cases and controls. If treatments are assigned in a truly random manner then the cases and controls should be equivalent in all respects apart from the treatment itself.

Randomized controlled experiments provide a gold standard for policy evaluation. Unfortunately, true experiments are nearly impossible (or at least highly impractical) in a policy setting because treatments cannot be randomly assigned. The denial of public support to those who otherwise qualify could lead to accusations of preferential treatment. Instead, legislative statute or administrative guidelines typically determine who qualifies for targeted development programs.

Quasi-experimental methods attempt to replicate key characteristics of a scientific experiment by measuring changes before and after a treatment (such as an training incentive or designation of an enterprise zone) and comparing these changes to that of a comparison group that did not receive the treatment. Such approaches are rarely used in a policy setting, because most programs do not collect detailed information about non-participants. Instead, many policy evaluation designs are based upon 'treatment-only' designs, where post-program impacts are compared to pre-program levels. While common, treatment-only designs are subject to a number of inferential biases – most notably their inability to account for the influence of external events occurring between the pre-test and post-test (i.e. history) and/or natural increases or decreases in outcome levels over time (maturation) (Cook and Campbell 1979). Interrupted time-series designs that include multiple observation points before and after a treatment can help control for some problems associated with time, but they still do not fully satisfy the 'what-if' counterfactual.

---

[1] All of the analysis in this study requiring direct access to confidential data was conducted on site at the CWIR offices in Augusta, Maine. Analysis conducted offsite was based on aggregate statistics that were reviewed and approved by CWIR staff for prior to release.

The addition of a comparison group of non-treated observations can greatly reduce the potential bias of treatment-only designs by offering a counterfactual scenario that can serve as a baseline for measuring program impacts. However, a comparison group, in and of itself, does not guarantee probabilistic equivalence. If program participants differ in systematic ways from those that did not participate, a comparison of participants to non-participants can either overstate or understate program impacts. This creates a serious evaluation problem known as selection bias. For example, some economic and workforce development programs target assistance to businesses that are growing or have strong growth potential (Bartik 2002). If these pre-destined 'winners' are more likely to receive treatment than a comparison group of other businesses, selection bias would likely overstate the beneficial impacts of the program. Conversely, some place-based development initiatives (such as state enterprise zones) target impoverished areas by providing businesses that move to these zones with favorable tax breaks, training, and technical assistance (Bondonio and Engberg 2000; Peters and Fisher 2002; Elvery 2009). Under this scenario, selection bias may underestimate program impacts because the treated areas were already disadvantaged relative to other places that were not selected.

In the absence of randomization, a next best alternative for reducing selection bias is to identify the attributes of businesses or workers that correlate with their selection into the treatment group and then screen the comparison group based upon these characteristics. Traditional methods for identifying equivalent comparison groups typically involve stratifying non-treated observations so that they match the treated observations based on a relatively small number of pre-treatment attributes (Isserman and Merrifield 1982; Isserman and Rephann 1995).

Propensity score matching methods have recently emerged as a parsimonious alternative for identifying a suitable comparison group while simultaneously accounting for a potentially large number of attributes (Rosenbaum and Rubin 1983; 1985; Heckman and Hotz 1989; Heckman *et al.* 1997; Heckman *et al.* 1998). Propensity score matching follows a two-stage process. In the first stage, it uses a probability model (such as a probit or logit) to estimate the probability of receiving treatment conditional on a set of covariates that may explain or correlate with program participation. The predicted probabilities from this model condense the relevant selection criteria into a single (propensity) score. In the second stage, a comparison group with similar propensity scores to the treatment cases is drawn from the non-treatment population.

Massive data requirements of the propensity score approach present a formidable barrier to its widespread adoption for economic development policy evaluation. The propensity score approach assumes that there is sufficient overlap between the sets of treatments and non-treatments. If self-selection in the treatment group is pervasive, it may not be possible to find equivalent matches from the non-treatment group. As such, propensity score matching works best when there is a large pool of non-treatment observations from which to choose comparison cases. The propensity score matching method also assumes that the process governing selection into the treatment group can be adequately modeled with existing data. A related complication is that businesses and workers in different industries are not subject to the same external pressures from international competition, technological change, business

cycles, or changes in local economic conditions. To fully control for these external influences, one must carefully select a comparison group that faces similar circumstances to the treatment group, preferably from the same or a closely related industry.

Confidential database collected through state unemployment insurance programs hold considerable promise for more rigorous and objective approaches to program evaluation. Unemployment insurance databases report data for the most disaggregate economic units possible—establishments and workers. These establishments and workers can be matched to administrative records of economic and workforce program participants to distinguish treatment and non-treatment groups. These confidential databases are also very large, offering a near comprehensive census of the private sector economy, increasing the likelihood of finding suitable matches. These databases are also reported at frequent and regular intervals (quarterly), which can be matched across successive periods to create a longitudinal record for each establishment and worker, facilitating before-after comparisons that can control for history and maturation bias. The next section describes these databases and discusses the process for building a longitudinal, matched employer and employee database for measuring program impacts.

Building a matched employer-employee database

This study follows the definition of the Maine Marine Trades cluster used by the NSAI for identifying qualifying businesses. As defined by the NSAI, the Marine Trades cluster includes boat builders, composite materials fabricators, marinas and boat yards, sail makers and canvas shops, rigging and cordage manufacturers and wholesalers, marine diesel engine mechanics, naval architects, servicing dealerships, marine supply outfitters, yacht brokerages, and other related businesses. The cluster also includes other regional assets and institutional actors (such as research centers and industry associations) that, while not part of the private sector economy, nonetheless provide important support and service to the cluster.[2] This study focuses soley on outcomes for private sector employers and their workers.

I identify companies in the Marine Trade cluster using a statewide inventory of Marine Trades companies developed by the NSAI to aid in their marketing, outreach, and program evaluation efforts. To populate this inventory the NSAI reviewed industry association membership lists, distributor client lists, advertisements in trade publications, discussions with knowledgeable industry representatives, and through a survey of owners and managers of Marine Trades companies who were asked to identify key suppliers, customers, competitors, sub-contractors and other business partners. The NSAI registry does not include information on employment, wages, sales, or other possible measures of performance outcomes. To develop a database of employment and wage outcomes, I matched businesses in the NSAI database to a statewide businesses registry known as the Quarterly Census of Employment and Wages (QCEW). The QCEW is a confidential database maintained by the Maine Department of Labor's Center for Workforce Research and Information (CWIR). It covers all

---

[2] The NSAI identified the key components of the Marine Trades cluster by reviewing recent studies on Maine industry clusters (see Colgan *et al.*, 2002; Colgan *et al.*, 2008; and Feser *et al.* 2009); conducting focus groups with members of relevant trade associations; and consulting with other experts that understand the cluster and its key technologies, markets, and labor force requirements.

businesses in the state that are required to report under the unemployment insurance program. It does not include sole-proprietors and other self-employed persons that have no payroll employees and does not fully cover all agricultural or government employees. It includes quarterly establishment-level information on the number of payroll employees, payroll earnings, the establishment's primary industry (i.e. NAICS code), whether it is independent or part of a firm with multiple units or branches, and its physical and mailing addresses.

I use the QCEW to track employment change in Marine Trades-related companies from the first quarter of 2000 to the first quarter of 2009, the most recent quarter available. To do this it was first necessary to build a longitudinal file, matching each establishment across successive quarters according to their Employer ID (EIN) and reporting unit (RNUM) identifiers. I then matches each company in the NSAI registry to the longitudinal QCEW based on a hierarchical algorithm that matched companies according to a string search of the company's trade and/or legal name, its owner's name, and the company's physical address. Of the 732 Marine Trades companies listed in the NSAI database, all but 200 were successfully matched to the QCEW. The bulk of the unmatched businesses are sole-proprietors, which are not covered under the state unemployment insurance programs. I also excluded 80 additional companies, either because they had no employment and wages on or after the first quarter of 2007, had anomalous (and likely erroneous) employment and wage entries, or were not private sector businesses (e.g. training providers, non-profits, and industry associations).

The second major confidential data source used in this study is the quarterly Unemployment Insurance (UI) database. Like the QCEW, it covers all payroll workers employed by businesses covered by the state's Unemployment Insurance program and does not include self-employed persons, informal, or otherwise non-payroll workers.[3] The UI database includes a unique ID number for each worker, EINs for each worker's employers during the quarter, and each worker's quarterly payroll earnings for each employer. I use the EIN numbers to establish a link between the workers in UI and their employer's attributes in the QCEW. From this matched employer-employee database, I identify workers who were employed by companies in the Marine Trades cluster between 2007 and 2009. I also track these workers backward in time by matching preceding quarters of the matched UI/QCEW database by each workers unique ID number. This allows me to develop an employment history for each worker that I use to measure changes in their employment status and wages across, regardless of their employer or industry.

The final step in this process is to identify businesses and workers who participated in NSAI training programs. The NSAI provided a separate database of all workers and businesses that participated in its training programs. This database contains, among other things: ID numbers for all workers as listed in their UI records; company names and EIN numbers; approximate training start dates; the amount of NSAI funds expended to sponsor the training; and a variety of other program and participant information.

---

[3] Although the QCEW and the UI database cover the same universe, the two databases differ in several key respects. The primary difference is while the QCEW measures jobs (i.e. employment) the UI database measures workers. The number of workers measured in the UI database uniformly exceeds the number of jobs reported in the QCEW, because during any given quarter a single job may shift between more than one worker.

I matched each trainee and business in the NSAI program database to their corresponding entries in the merged UI/QCEW database, denoting the year and quarter when NSAI training began. In the rare cases where the worker was associated with multiple employers during a single quarter, I use NSAI program data to identify the employer whom actually supported the training. According to the NSAI database, there were 794 trainees and 85 companies that participated in NSAI training programs by the end of 2008. Of these, 575 trainees and 43 businesses were successfully identified in the QCEW-UI database and are included in this analysis as the 'treatment' groups.

Employment impacts of NSAI training programs
This study does not provide a comprehensive assessment of the impacts of NSAI training programs and activities. Instead, I focus narrowly on two questions of common interest to economic development policy and of specific interest to economic development objectives of the NSAI:

1. Did establishments participating in NSAI training programs add more jobs than similar companies that did not participation (i.e. job creation)?
2. Did trainees earn higher wages following NSAI training compared to similar workers who did not receive training (i.e. wage growth)?

There is some theoretical basis to justify training subsidies as a source of job creation and wage growth, at least in the short term. Training subsidies constitute a direct subsidy to labor training which may result in more jobs as firms shift to more labor-intensive forms of production. The company may also use the subsidy to offset a wage increase, especially if the training results in more valuable and productive employees. However, the NSAI subsidies are not permanent, nor are they very large. For example, under the typical NSAI trainee received a direct subsidy of roughly $2,000 for their employer – considerably less than the typical Marine Trades worker's annual payroll earnings of roughly $27,000. Similarly, the average subsidy per business amounts to just over $10,000 – a non-trivial amount, but still a mere fraction of the total wage bill for most firms.[4] As such, even if such programs do have employment and wage impacts, I would not expect them to be large.

The remainder of this section addresses the first question by comparing the before and after change in employment among establishments participating in NSAI training programs to Marine Trades companies that did not participate. To evaluate the possible influence of selection bias, I conduct my analysis for two separate comparison groups; (1) the full set of non-participating Marine Trades establishments; and (2) a screened sample of non-participating establishments that have been matched to the treatment group through propensity scoring.

For the first stage of the propensity score method, I use a logistic regression to estimate the probability of participation in NSAI programs based purely upon their pre-NSAI attributes.

---

[4] In addition to the subsidies for businesses and workers, the NSAI also covered many of the direct and indirect costs of the training, such as curricula development costs, materials costs, instructor training, facilities fees, etc.

All of these measures were either taken directly from or derived from information in the longitudinal QCEW. These variables are:

- A dummy variable indicating whether the establishment was born after the 1st quarter of 2000.
- Size as measured by the establishment's employment in the 4th quarter of 2006.
- Recent employment growth, measured by change in employment between the 4th quarter of 2006 and the 4th quarter of 2005.
- Each company's quarterly payroll earnings per worker, measured in the 4th quarter of 2006.
- Recent earnings growth the company, measured as the growth rate of the firm's annual per worker payroll calculated as an annual average from 2004 to 2006.
- Dummy variable used to distinguish boat yards (BY) and boat builders (BY) from other companies in the Marine Trades cluster.

I also include second order terms for the continuously measured variables to accommodate possible nonlinearities.

While presented mainly as an intermediate step in estimating program impacts, the first-stage regression provides some interesting insights into the types of companies selected for participation in NSAI training programs. The coefficient estimates produced from the model are not directly interpretable, although they can be easily converted into odds ratios by taking the exponent of each coefficient estimate. The results of the establishment-level logistic regression are provided in Table 1. There were 74 treatment and 311 non-treatment observations included in the first stage logit. The remaining companies had missing data in one or more of the independent variables, and subsequently are not included in the remainder of the analysis. The generalized R^2 value of .20 and significant likelihood ratio statistic indicate a generally good fit, considering the likely complexity of participation decisions and the limited number of variable available.

[Table 1 near here]

Establishment size, growth, and salary levels are all significant and positively associated with NSAI participation. The squared terms for these coefficients are also significant, suggesting possible diminishing returns. Larger companies typically have greater administrative resources to devote to access public resources and, because they are larger, are more likely to have had previous contact with state and local economic and workforce development agencies. Growing companies might also more need for training, especially for their new employees. Higher salaries reflect an investment in workers, and the company may use training as a complementary strategy for retaining valued employees.   Boat builders and boat yards are also more likely to participate, commiserate with the predominantly production-oriented training programs of the NSAI.

The second stage of the propensity score approach uses an algorithm to identify a comparison group with similar propensity scores to participating companies. There are many possible

matching algorithms and strategies, ranging from nearest neighbor matching, to stratification, to kernel weighted group matching.[5] I chose a caliper-matching approach with one-to-one non-replacement matching for this study. Without delving too far into the technical details, in one-to-one caliper matching the analyst defines a probability threshold (i.e. support region) for each treatment, and selects the closest neighboring non-treatment case within this threshold. I use a fairly stringent caliper setting of .1. While this may result in some non-matched treatments if no non-treatments fall within the support region, it also reduces bias from inappropriate matching. The one-to-one matching also produces a balanced set of treatments and controls and is relatively easy to explain to a policy audience, at least in comparison to more elaborate kernel-weighting approaches.

Table 2 illustrates the value of the propensity method in addressing selectivity bias by comparing key attributes of participating and non-participating establishments, before and after propensity matching. I matched fifty-eight of the original 77 participating companies. Before matching, participating firms were significantly larger and paid higher salaries than the typical non-participant. Recent employment growth and payroll growth did not differ significantly between participating and non-participating companies. After matching, there were no significant differences between treatment and control groups.

[Table 2 near here]

I estimate the average treatment effect of NSAI training on establishment job creation using a difference-in-differences estimator. For each establishment, I calculate the mean difference in employment before and after NSAI training and compare this to the mean difference experienced by the comparison group. To control for seasonal fluctuations, the before and after measurement periods must be spaced one year apart. I provide separate estimates for immediate impacts and short-term effects. I measure immediate impacts by comparing the post-NSAI employment levels in the quarter immediately following training to the pre-training employment levels one year prior. For example, for a company first participating in NSAI training in the 1st quarter of 2007, I compare their employment in the 2nd quarter of 2007 (one quarter after the training) to the 2nd quarter of 2006 (one year differences). For the 33 establishments who participated in NSAI programs on or before the first quarter of 2008, I also test for average treatment effects over a slightly longer period (i.e. short-term effects) by measuring employment change for one year before training to one year after. The results are reported in Table 3.

[Table 3 near here]

I find no evidence that NSAI training induced an immediate increase in employment. However, NSAI training may have had some beneficial influence on establishment employment levels over a slightly longer period. Three quarters prior to NSAI training, the treatment group of program participants employed, on average, 54 workers. One quarter after training, average employment dropped slightly to 52 workers per establishment. Employment levels among the comparison group fell by a similar amount. Among the 36 establishments for

---

[5] For a comparison of alternate matching strategies in the context of training program evaluation see Mueser *et al.* (2007).

which a full-year of post-treatment effects could be measured, average employment levels increased very slightly, from 23 to 24 employees per firm. During this same period, employment in the comparison group declined by approximately 3 workers per establishment. The resulting difference in difference average of 3.81 employees was statistically significant at the 95 percent level using a single tailed t-test.

Wage impacts of NSAI training programs

In this section, I use a similar approach to test whether workers receiving NSAI-supported training experienced a short-term impact in their quarterly earnings compared to group of similar Marine Trades workers that did not.

In the first-stage logistic regression, I model the probability of an employee's participation in NSAI training programs based upon both employee and employer characteristics, respectively derived from the UI and QCEW databases. In addition to the characteristics used in the establishment-level models, I also condition worker participation in NSAI programs upon the each worker's average annual payroll earnings in 2006, the annual average growth of their earnings from 2004 to 2006, their length of time with their current employer, and their length of time in the Maine labor market. I also include squared values for each of the continuous variables.

Four hundred and thirty eight trainees and 6046 non-trainees were included in the first-stage logistic regression (Table 4). An increase in the length of time with their current employer has a negative association with participation, meaning that newer employees are more likely to receive training compared to long-time employees. The typical NSAI trainee has also had slightly higher earnings in the fourth quarter of 2006 and faster earnings growth from 2004 to 2006. Similar to the establishment-level model, participating workers also tend to work for boat builders and companies that tend to pay higher wages. Contrary to the establishment model, the average establishment size for participating workers is slightly smaller than the average size of firms whose workers did not participate. At first appearance, the results on size may seem at odds with earlier findings that larger firms were more likely to participate. This may be because the establishment model does not consider for the number of trainees that work for the same employer. If most of the trainees come from a relatively small number of smaller firms, the observed average firm size of participants may be smaller worker-level model, but larger in the establishment-level model.[6]

[Table 4 near here]

Similar to before, I use a caliper-based matching algorithm to pair participants with non-participants based upon the predicted probabilities from the first stage regression. Table 5 provides a side-by-side comparison of NSAI trainees to non-trainees based upon their pre-NSAI training characteristics (i.e. prior to 2007), with and without the propensity score

---

[6] An alternate, and possibly more appropriate, specification uses a multi-stage model to first estimate the probability of employer selection, and then estimate the probability of worker's participation given that their employer was selected. In a recent study, Elvery (2009) used a similar framework to evaluate worker outcomes from enterprise zone designations. This approach will be explored in future work.

matching. NSAI trainees clearly differ from the full set of non-trainees. NSAI trainees have, on average, spent less time in the Maine labor force and less time with their current employer. Prior to NSAI training, they had significantly higher earnings growth than non-trainees. Trainees also tended to work for smaller firms that paid slightly higher wages than the employers of non-trainees. While the employers of trainees tended to add fewer workers from 2005 to 2006, they had a significantly higher two-year employment growth rate. Business payroll growth was not found to significantly differ between trainees and non-trainees.

[Table 5 near here]

After the propensity score matching, the comparison set of non-trainees was reduced from 7,495 to 423, but this smaller set is a much closer match to the trainees. The only remaining significant differences between trainees and non-trainees were in employer characteristics: the employer's employment growth from 2005 to 2006, their two-year average employment growth rate, and the employer's average payroll per worker. While still significant, the magnitude of the differences in employment growth and the employment growth rate are much less than before matching. Although the average earnings gap between companies of trainees and non-trainees had widened slightly.

The difference-in-differences estimates of the average treatment effects show no significant immediate increase in trainee quarterly wages (Table 6). However, there is some evidence of significant differences measured over a slightly longer period. In the quarter immediately following NSAI training, the real average quarterly earnings of program participants was $124 lower compared to the year before, while the real earnings of non-participants was nearly unchanged, declining by only $15. However, the $110 post-pre difference between the treatment and comparison groups is not statistically significant. In the two year period before and after NSAI training, the real quarterly earnings of non-trainees declined by $130, while the trainee earnings grew by $248. The $378 relative increase for trainees is statistically significant at the 95% level for a one-tailed distribution.

[Table 6 near here]

Summary, Limitations and Next Steps
In this paper, I have discussed the value of combining quasi-experimental methods with employee- and establishment-level labor force data for evaluating workforce and economic development initiatives. These methods help address some of the persistent difficulties in evaluating policy. More empirically rigorous approaches to evaluation can help offset the difficulties posed by the lack of objective data collection or outcome measures; the need to distinguish program impacts from business cycles and market fluctuations; and bias from participant selectivity that may cause program impacts appear either larger or smaller than they truly are.

I demonstrate these methods by evaluating the short-term impacts of North Star Alliance Initiative (NSAI) training programs on employee earnings and on the employment growth of participating establishments. The results suggest that NSAI programs may have had a small

positive influence on employment levels of participating establishments, and on the wage levels of participating workers. The influence of the program is not immediate, differences measures in the first quarter after training were not significant and of the wrong direction. The differences were significant one year after training, but still small in magnitude. In the case of both employment and wages, the significant differences between treatment and comparison cases was due to a combination of a slight positive increase for participants coupled with a modest decline in employment and earnings for non-participants. Given the growing economic recession of 2008 and 2009, perhaps a more accurate interpretation of the employment and wage impacts of NSAI programs is they enabled participants to better weather the recession relative to non-participants.

The use of micro-level data in a quasi-experimental design is a more rigorous and objective approach to program evaluation that is commonly used. However, this study still has some limitations. It is difficult to distinguish short-term fluctuations from long-term trends with only a few periods of post-treatment data. At best, this study can only shed light on the short-term impacts on NSAI workforce development activities. The impacts of economic development, training, and education are particularly difficult to assess over the short-term, because the benefits to education are cumulative and accrue over a longer time horizon -- sometimes over a lifetime.

The broader impacts from training are also difficult to codify and measure, and cannot be fully captured by an approach that only examines individual- and establishment-level outcomes. Training subsidies may provide a short-term boost to wages and induce more hiring, but the real benefit of training to long-term economic development is through the influence of enhanced human capital on a company's productivity and their ability to adapt to new production methods. Over the long-term, we would expect training subsidies to increase the productivity of participating businesses making them more profitable and likely to survive and expand compared to their counterparts that do not invest in their workforce. Training may also help to reduce the long-term costs to business from employee turnover by investing in their employee's career development. The entire state economy may also benefit from the accumulation of human capital as additional businesses gravitate to places with deep pools of skilled workers. However, for training programs to have broader economic development impacts they must be sustained and continually adapted to ever-changing technological and market conditions. For this reason, the reader is advised to interpret the outcomes of this study as only partial and preliminary. A more detailed study covering a broader range of outcomes conducted over a longer time-span is necessary to evaluate the full economic impacts of training and other types of endogenous development strategies.

The seasonal nature of hours and employment in the Marine trades industry further complicates the analysis. I found considerable volatility in the quarter-to-quarter earnings of individual workers and strong seasonal variations in employment. In this preliminary study, I attempt to account for seasonality by comparing pre- and post-treatment wages and employment at similar points in the annual cycle. A superior approach would take full advantage of the time-series dimension of the longitudinal databases to statistically control for seasonal volatility and distinguish immediate pulse effects from trend effects. As this work

progresses, I plan to measure program impact using an integrated time-series case-comparison approach, such as a repeated-measures ANOVA or similar type of interrupted time-series analysis.

Although they constitute a considerable improvement over existing aggregate secondary data sources, the UI and QCEW also have some limitations. Neither the was created with the intent of evaluating government initiatives, and both lack some of the key variables necessary to definitively estimate impacts or account for all of the characteristics associated with selection. For instance, the UI database only shows each worker's gross quarterly earnings, not the number of hours worked. Therefore, a raise in an employee's hourly wage rate may or may not translate into higher quarterly earnings in the UI if the increase in wages is offset by a decline in hours. Similarly, the UI and QCEW only contain a limited number of employee and establishment characteristics. Missing are important demographic characteristics of workers, such as gender and race, more specific data on the skill set and job tasks, and relevant establishment characteristics such as the types of technologies used in production or proximity to training programs. If these and other missing characteristics are correlated with participation in NSAI programs, the matching set may not fully account for selection bias. While some of this data could potentially be collected for participants through the application process, it cannot be as easily collected on non-participants in the absence of widespread surveying.

Both the UI and the QCEW require considerable processing and quality control before they can be released for analytical use. This creates a lag in the time between when the data is collected and when it released—usually within three-quarters to a year after its initial collection date. Compared to most publicly produced secondary data series, a year lag is actual a rather swift turn-around. However, in may not be quick enough to use this data for continual monitoring and benchmarking of on-going programs. It is much better suited for post-hoc evaluation and assessing the effectiveness of long-term policies and programs.

Although the use of quasi-experimental designs coupled with large-scale micro databases holds considerable promise for *reducing* the problem of selection bias in economic development policy research, in most realistic policy contexts they will not be able to entirely *eliminate* it. Propensity score matching is only effective in controlling for selection bias to the extent that the factors governing or correlated with participation can be accurately measured and modeled. Participation in economic development programs is complicated and relates a number of factors that are difficult to quantify. These may include the effectiveness of program outreach efforts, the timing of the company's production schedule in relation to course scheduling, the company's interest in skills development and new technologies, as well as the manager's general trust of government that may be conditioned upon either ideology or their previous experience with public initiatives and officials. Furthermore, to avoid circularity between cause and effect, the comparison group is screened by their pre-treatment attributes. Companies that *anticipate* growth in the near future are likely to view training more favorably and also experience higher levels of employment and wage growth. Controlling for employment and wage growth may account for some of this bias, but not all.

Works Cited

Bartik, T. J. (2002). Evaluating the impacts of local economic development policies on local economic outcomes: What has been done and what is doable? Upjohn Institute Staff Working Paper No. 03-89. Kalamazoo, MI, W.E. Upjohn Institute for Employment Research: 40.

Bondonio, D. and J. Engberg (2000). "Enterprise zones and local employment: Evidence from the states' programs." Regional Science and Urban Economics 30: 519-549.

Colgan, C. S. and B. H. Andrews (2008). Maine's Technology Sectors and Clusters: Status and Strategy. Augusta, ME, Office of Innovation Maine Department of Economic and Community Development.

Colgan, C. S., C. Baker, B. Nan and M. Cote (2002). Assessing Maine's Technology Clusters. Hallowell, ME, Maine Science and Technology Foundation.

Cook, T. D. and D. T. Campbell (1979). Quasi-Experimentation; Design & Analysis for Field Settings. Boston, Houghton Mifflin Company.

Elvery, J. A. (2009). "The Impact of Enterprise Zones on Resident Employment: An Evaluation of the Enterprise Zone Programs of California and Florida." Economic Development Quarterly 23(1): 44-59.

Feser, E., H. C. Renski and J. Koo (2009). Regional Cluster Analysis with Interindustry Benchmarks. Targeting Regional Economic Development (TRED). Goetz, S., S.C. Deller and T. Harris (eds.).

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998). "Characterizing Selection Bias Using Experimental Data." NBER Working Papers: 6699.

Heckman, J. J. and V. J. Hotz (1989). "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." Journal of the American Statistical Association 84(408): 862-874.

Heckman, J. J., H. Ichimura and P. E. Todd (1997). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme." Review of Economic Studies 64(4): 605-654.

Isserman, A. and T. Rephann (1995). "The economic effects of the Appalachian Regional Commission: An empirical assessment of 26 years of regional development planning." APA Journal 61(3): 345-364.

Isserman, A. M. and J. Merrifield (1982). "The use of control groups in evaluating regional economic policy." Regional Science and Urban Economics 12(1): 43-58.

Mueser, P. R., K. R. Troske and A. Gorislavsky (2007). "Using state administrative data to measure program performance." Review of Economics and Statistics 89(4): 761-783.

Peters, A. H. and P. S. Fisher (2002). State Enterprise Zone Programs: Have They Worked?, Kalamazoo, Mich.: W. E. Upjohn Institute for Employment Research.

Planning Decisions, L. (2007). Maine's Boat Building Industry: Obstacles & Opportunities. South Portland, ME, report prepared on behalf of Maine Built Boats.

Rosenbaum, P. R. and D. B. Rubin (1985). "Constructing a control group using multivariate matched sampling models that incorporate the propensity score." American Statistician 39(1): 33.

Rosenbaum, P. R. and D. R. Rubin (1983). "The central role of the propensity score in observational studies for causal effects." Biometrika 70(1): 41-55.