The Effects of Vocational Rehabilitation for People with Mental Illness

David Dean
University of Richmond

John Pepper
University of Virginia

Robert Schmidt
University of Richmond

Steven Stern*
University of Virginia

January 2014

Abstract

We construct a structural model of participation in vocational rehabilitation for people with mental illness. There are multiple services to choose among, and each has differerent effects on employment, earnings, and receipt of SSI/DI. We estimate large effects for most of the services implying large rates of return to vocational rehabilitation.

1 Introduction

The public-sector Vocational Rehabilitation (VR) program is a \$3 billion federal-state partnership designed to provide employment-related assistance to persons with disabilities. While thought to play an important role in helping persons with disabilities to engage in gainful employment and possibly reducing disability insurance roles (Loprest, 2007), very little is known about the long term-efficacy of VR in the United States. The last published economic evaluation of the U.S. public-sector VR program is from over 10 years ago (Dean, Dolan, and Schmidt,1999).[1]

[1] Although certainly informative, the earlier studies have a number of methodological shortcomings and have only limited relevance to the current VR system which serves a clientele with a much wider range of impairments. Other early evaluations include Conley (1969); Bel-

In this paper, we study the impact of the VR program using a unique panel data source on all persons who applied for services in the state of Virginia in State Fiscal Year 2000. Combining these data with a structural model of service provision, we are able to estimate the long-term impact of VR services on employment, earnings, and disability insurance receipt. These results are then used to simulate the distribution of rates-of-return on VR services.

Given that the impact of VR services is thought to differ by the type of limitation (Dean and Dolan,1991; Baldwin 1999; Dean, Dolan, and Schmidt, 1999; and Marcotte, Wilxox-Gok, and Redmond, 2000), we focus on VR clients diagnosed with mental illness, an increasingly important part of the VR case-load. Originally established in 1919 to provide restorative services to persons with primarily physical disabilities, the program's emphasis has shifted in recent decades to serve persons with cognitive impairments or mental illness. While comprising an ever-larger share of the VR clientele, the latter group has turned out to be particularly hard to serve. As the Government Accountability Office (2005) notes, persons classified with mental or psycho-social impairments make up almost one-third of VR program exiters nationwide in 2003 but, at 30%, had the lowest employment rate outcome of all groups served.[2] Consequently, an increasing share of VR expenditure, along with research and practice in the VR and mental health fields, has been concentrated on increasing the employability of persons with mental health problems.[3]

Importantly, our administrative data from the 2000 applicant cohort in Virginia is much richer than that used in previous analyses. Other economic analyses of VR efficacy (see Conley, 1969; Bellante, 1972; Worrall, 1978; Nowak, 1983) have relied almost exclusively on the Rehabilitation Service Administration's RSA-911 Case Service Report of nationwide closures from the VR program. The problems with evaluations based on these RSA-911 data are manifold. First, a censoring problem arises because the RSA-911 sample frame is drawn from cohorts of cases terminated from the program during the same

---

lante, (1972); Worrall (1978); Berkowitz (1988); Dean and Dolan (1991). Several more recent studies evaluate the European active labor market programs for persons with disabilities (e.g., Raum and Torp, 2001; Bratberg, Grasdal, and Risa, 2002; Frolich, Heshmati, and Lechner, 2004; Aakvik, Heckman, and Vytlacil, 2005; and Bewley, Dorset, and Salis, 2007).

[2] Kessler et al. (2001) estimates that more than 25% of U.S. adults had a mental illness in the previous year, with 7% having a major depressive disorder and 18% having anxiety disorders. The prevalence of mental illness among adults in the United States imposes severe employment consequences with unemployment rates for persons with severe mental illness estimated to be as high as 95% (Mueser, Salyers, and Mueser, 2001).

[3] The increased emphasis on achieving competitive employment outcomes for persons with mental health disorders has led to numerous studies published in the VR literature that examine specific interventions for persons with varying degrees of mental illness. See Bond, Drake, and Becker (2001) for a review of such analyses or Cook et al. (2005); Burns et al. (2007); or Campbell et al. (2010) for descriptions of specific experiments. These investigations typically consist of small clinical trials of a specific intervention of supported employment versus the more traditional VR practice of .train and place.. Such randomized clinical studies are typically of short duration and thus lack sufficient information on longer-term employment outcomes. Ultimately, this type of analysis is not suited for evaluating the on-going VR program, which legally is not allowed to engage in randomized control studies using federal support.

2

year. This is a significant drawback for a program with a wide variation in program duration that results in comparing cohorts who applied for services over different time periods. By focusing on an applicant cohort, we avoid this censoring problem. Second, the RSA-911 reports earnings only at two points: 1) self-reported weekly earnings at the time of referral to the VR program and 2) following three months of employment if employed. As Loprest (2007) notes, these analyses suffered from the RSA-911's lack of longitudinal earnings. In our data, we observe quarterly employment and earnings data as well as VR service data from 1995 to 2008. Thus, using data on individual quarterly employment and earnings prior to, during, and after service receipt, we examine both the short- and long-term effects of VR services. Finally, evaluations using the RSA data classify clients as either receiving or not receiving substantial VR services. In practice, however, VR agencies provide a wide range of different services which are likely to have very different labor market effects. Using the administrative data from Virginia, we examine the impact of specific types of services rather than just a single treatment indicator. In particular, following Dean et al. (2002), we aggregate VR services into six types – diagnosis and evaluation, training, education, restoration, maintenance and other – and allow these six services to have different labor market effects.

Another important contribution afforded by the richness of our data is that we evaluate the impact of VR services on the receipt of payments from the Social Security Administration's Disability Insurance (DI) and Supplemental Security Income (SSI) programs. As the enrollment and costs of disability insurance programs have grown over the past two decades, there has been growing interest in whether VR programs might serve to reduce the number of persons receiving DI/SSI benefits (e.g., Autor and Duggan, 2010; Stapleton and Marin, 2012). This is especially true for persons with mental illness who constitute the largest and most rapidly expanding subgroup of DI/SSI program beneficiaries (Drake et al., 2009). If VR services improve labor market outcomes of potential DI/SSI beneficiaries, some clients may choose to fully participate in the labor market rather than take up DI/SSI. Yet, VR programs may instead lead to an increase in take-up by serving to help clients understand the DI/SSI programs and rules (Stapleton and Martin, 2012). Although there are a handful of studies assessing the correlation between VR services and DI/SSI receipt (e.g., Hennessey and Mueller, 1995; Tremblay et al., 2006; Rogers, Bishop, and Crystal, 2005; and Stapleton and Erickson, 2004), research on the impact of VR on DI/SSI receipt is limited (Stapleton and Marin, 2012).

Finally, we formalize and estimate a structural model of endogenous service provision and labor market outcomes. Except for controlling for observed covariates, the existing literature does not address the selection problem that arises if unobserved factors associated with VR service receipt are correlated with latent labor market outcomes. Hotz (1992) provides a framework for the Governmental Accountability Office that laid out several options for evaluation of the public-sector VR program in a non-experimental setting that included both parametric and non-parametric techniques to control for the problem of selection bias inherent in such voluntary programs. Although several studies

of the European active labor market programs for persons with disabilities incorporated such methodologies (e.g., Raum and Torp, 2001; Bratberg, Grasdal, and Risa, 2002; Frolich, Heshmati, and Lechner, 2004; and Aakvik, Heckman, and Vytlacil, 2005), evaluations of VR programs in the U.S. have not kept up with the significant advances made during the past two decades in evaluations of manpower training programs (see, for example, Imbens and Wooldridge, 2009).[4] We address the selection problem using instrumental variables that are assumed to impact service receipt but not the latent labor market outcomes, pre-program labor market outcomes that control for differences between those who will and will not receive services, and a formal structural model of the selection process.

The paper proceeds as follows: Section 2 describes the economic model used throughout the paper. We construct a multivariate discrete choice model for service provision choices. We augment that with a probit-like employment equation and an earnings equation. We allow for correlation of errors among all of the equations. Next, we describe the three sources of data used in our analysis in Section 3 and the econometric methodology used to estimate the model from Section 2 in Section 4. Estimation results are presented in Section 5, and a rate-of-return analysis is presented in Section 6. Our results imply generally high rates of return but with significant variation in returns across people with varying characteristics. We also find the VR services increase the probability of DI/SSI receipt.

2 Model

Let y_{ij}^* be the value for individual i of participating in VR service j, $j = 1, 2, .., J$, and define $y_{ij} = 1\left(y_{ij}^* > 0\right)$ be an indicator for whether i receives service j. Assume that

$$\begin{aligned} y_{ij}^* &= X_i^y \beta_j + u_{ij}^y + \varepsilon_{ij}, \\ \varepsilon_{ij} &\sim Logistic \end{aligned} \tag{1}$$

where X_i^y is a vector of exogenous explanatory variables, and u_{ij}^y is an error whose structure is specified below.

Next, we introduce three equations associated with the value of working, log-quarterly earnings, and the value of received DI/SSI payments. Let z_{it}^* be the value to i of working in quarter t, and define $z_{it} = 1\left(z_{it}^* > 0\right)$. Assume that

$$z_{it}^* = X_{it}^z \gamma + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^z y_{ij} + u_{it}^z + v_{it}^z \tag{2}$$

---

[4] Dean and Dolan (1991) follow advances in the more general field of manpower training evaluation at the time (see, for example, Ashenfelter, 1978; Bassi, 1984; and Heckman and Hotz, 1989), but do not address the problem of selection on unobservables. Selection is thought to be a central problem in addressing the impact of job training programs (Card and Sullivan, 1988; LaLonde, 1995; Friedlander, Greenberg, and Robins, 1997; and Imbens and Wooldridge, 2009). Aakvik, Heckman, and Vytlacil (2005) find that that this selection problem plays an important role in the evaluation of a Norwegian VR training program.

where $X_{it}^{\tilde{z}}$ is a vector of (possibly) time-varying, exogenous explanatory variables, d_{ik} is a dummy variable equal to one iff the amount of time between the last quarter of service receipt and t is between time nodes τ_k and τ_{k+1},[5] and u_{it}^z is an error whose structure is specified below. The time periods implied by the nodes we use are a) 2 or more quarters before service onset, b) 1 quarter before service onset, c) 1 quarter after service onset to 8 quarters after service onset, and d) 9 or more quarters after service onset. Next let w_{it} be the log quarterly earnings of i at t, and assume that

$$w_{it} = X_{it}^w \delta + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^w y_{ij} + u_{it}^w + v_{it}^w \tag{3}$$

where variables are defined analogously to equation (2). Next, let r_{it}^* be the value to i of receiving SSI or DI payments in quarter t, and define $r_{it} = 1\left(r_{it}^* > 0\right)$. Assume that

$$r_{it}^* = X_{it}^r \psi + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^r y_{ij} + u_{it}^r + v_{it}^r \tag{4}$$

where variables also are defined analogously to equation (2).[6] This is the first paper to jointly model VR services, employment outcomes, and DI/SSI receipt.

Finally, assume that

$$
\begin{aligned}
u_{ij}^y &= \lambda_{j1}^y e_{i1} + \lambda_{j2}^y e_{i2}, & (5)\\
u_{it}^z &= \lambda_1^z e_{i1} + \lambda_2^z e_{i2} + \eta_{it}^z,\\
u_{it}^w &= \lambda_1^w e_{i1} + \lambda_2^w e_{i2} + \eta_{it}^w,\\
u_{it}^r &= \lambda_1^r e_{i1} + \lambda_2^r e_{i2} + \eta_{it}^r,\\
\eta_{it}^z &= \rho_\eta \eta_{it-1}^z + \zeta_{it}^z,\\
\eta_{it}^w &= \rho_\eta \eta_{it-1}^w + \zeta_{it}^w,\\
\eta_{it}^r &= \rho_\eta \eta_{it-1}^r + \zeta_{it}^r,\\
\begin{pmatrix} \zeta_{it}^z \\ \zeta_{it}^w \\ \zeta_{it}^r \end{pmatrix} &\sim iidN\left[0, \Omega_\zeta\right],\\
\begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix} &\sim iidN\left[0, I\right],\\
v_{it}^z &\sim iidN\left[0, 1\right],\\
v_{it}^w &\sim iidN\left[0, \sigma_w^2\right], \text{ and}\\
v_{it}^r &\sim iidN\left[0, 1\right].
\end{aligned}
$$

[5] In effect, we allow for level spline effects for service effects on labor market outcomes.

[6] The specification in equation (4) ignores all of the issues associated with actually applying for and being awarded disability benefits (e.g., see Kreider, 1998, 1999; Benitez-Silva et al., 1999; French and Song, 2012) or controlling for measurement error in disability and its interaction with disability benefits (e.g., see Benitez-Silva et al., 1999) because these are not the focus of this work.

We include the (e_{i1}, e_{i2}) to allow for two common factors affecting all dependent variables with factor loadings $\left(\lambda_{jk}^y, \lambda_k^z, \lambda_k^w, \lambda_k^r \right)_{k=1}^2$. We also allow for serial correlation and contemporaneous correlation in the labor market errors $(\eta_{it}^z, \eta_{it}^w, \eta_{it}^r)$. The covariance matrix implied by this error structure is presented in Appendix 8.1. See Dean et al. (2013a) for a similar structure applied to people with cognitive impairments.

3 Data

We use three main sources of data: a) the administrative records for the state fiscal year (SFY) 2000 applicant cohort of the Virginia Department of Aging and Rehabilitative Services (DARS), b) the quarterly administrative records on labor market activity of the Virginia Employment Commission (VEC) from 1995 to 2008 for those people in the DARS data, and c) the quarterly administrative records of the Social Security Administration on Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) benefit receipt from 1995 to 2008 for those people in the DARS data. We also merge these files with data from the Bureau of Economic Analysis on county-specific employment patterns. Each of these is discussed in turn below.

3.1 DARS Data

3.1.1 DARS Sample Frame

Our starting point is the administrative records of the Virginia DARS for the 10323 individuals who applied for vocational rehabilitative (VR) services in SFY 2000 (July 1, 1999 - June 30, 2000). Our analysis focuses on 1555 DARS clients with mental illnesses. We exclude individuals for the reasons specified in Table 1. The criterion associated with having a mental illness used for sample selection is that the primary or secondary diagnosis listed in the administrative records must be a mental illness in at least one quarter while the individual has an open case; this may be the first case in 2000, or it may be a subsequent case.[7] Not having a mental illness is the single most important reason for exclusion from our estimation sample, resulting in 6476 excluded observations. Because we need diagnoses for each case, we exclude 94 observations where primary and/or secondary diagnosis was missing as well. We also excluded 71 individuals with neither any service records nor employment records.[8]

---

[7] In those cases where the individual was not diagnosed in the first case but later was, we essentially are assuming that the individual had a mental health problem during the first case but it was not recorded. This might happen, for example, because there was some other, more dominant disability at the time.

[8] While it could be the case that such individuals applied to DARS and withdrew for some reason and were also never employed, we were concerned about including such observations because there was a reasonable chance of a problem with the merging of the DARS and VEC data. To the degree that we excluded valid observations, we are biasing our results toward finding no effect for DARS services because the excluded observations would have been

Table 1: Missing Value Analysis

Cause	# Obs Lost	Proportion of Total	# Remaining
Applicants in SFY 2000			10323
Missing or Questionable SSN	81	0.008	10242
Died While in Program	65	0.006	10177
Missing Gender or Date of Birth	1	0.000	10176
Not in Virginia	59	0.006	10117
Not Mentally Ill	6476	0.640	3641
Missing Primary Disability	87	0.024	3554
Missing Secondary Disability	7	0.002	3547
Initial Service Spell before SFY 2000	1220	0.344	2327
Age Younger than 21 Years	701	0.301	1626
Neither Service nor Employment Record	71	0.044	1555
Number Remaining in Sample	1555	0.151	

We focus on the "base case" defined as an individual's initial case in SFY 2000, recognizing that individuals can have multiple "service spells" or "cases," each of which includes an application and administrative closure. We have administrative information between SFY 1987 and 2007 that allows us to identify these multiple service spells and exclude observations where the individual's first service spell was prior to SFY 2000. We do this to avoid bias associated with left censoring (e.g., Heckman and Singer, 1984a). In particular, if the subsample of people who enroll in services more than once is different than those who enroll only once, then those people who had service spells prior to SFY 2000 will have unobservable characteristics different than those whose first spell is in SFY 2000. Dean et al. (2013a) find significant left-censoring biases for a sample of people with cognitive impairments.

3.1.2 DARS Data for Service Provision

Upon application, an individual's case is assigned to a counselor who assesses the individual's eligibility for the program. This assessment typically includes a diagnosis of the impairment. The case may be administratively closed at this point because the impairment is deemed insufficiently severe or too severe or because the individual withdraws from further consideration for VR eligibility. Beyond assessment and some counseling, these individuals receive few, if any, services.

By contrast, for those accepted for service, the counselor and individual develop an individualized plan for employment (IPE) which specifies the array of services to be provided. Services can include, for example, restorative medical care, vocational counseling and guidance, training (both vocational and rehabilitative), education, job search and placement, and/or assistive services. Some individuals drop out before completing the program, possibly having received

---

recorded as having no employment and no change in employment before and after service had we included them while, as can be seen in Figure .4, the average lifetime employment path displays declining employment.

little or no services beyond the development of an IPE.

Although the vast majority of vouchers provide the cost of the purchase service, some provide no information about cost. Thus, in Table 2 below, total purchased services (column 1) are disaggregated into those with recorded positive expenditures and those without (columns 2 and 3).

Services can be provided to an individual in any combination of three ways: a) as a "purchased service" through an outside vendor using DARS funds, b) as a "similar benefit" purchased or provided by another governmental agency or not-for-profit organization with no charge to DARS, and/or c) internally by DARS personnel. The DARS administrative records provide access to dates, quantities, costs, and types of purchased services. At least one voucher was recorded for 70% of base cases. Although the vast majority of vouchers provide the cost of the purchase service, some provide no information about cost. Thus, in Table 2, total purchased services (column 1) are disaggregated into those with recorded positive expenditures and those without (columns 2 and 3).

The DARS administrative data do not, however, reveal the same detailed information for in-house services or similar benefits. Instead, we measure non-purchased service provision using two additional sources of service information. First, DARS reports on the provision of similar benefits (but not timing or cost) for the Rehabilitation Service Administration RSA-911 Case Service Report due at the end of the federal fiscal year for all cases closed during that year. Use of this information is complicated by several factors, the most important being that the two indicators included for each service category sometimes provide inconsistent information. We impose the condition that this source identifies the provision of similar benefits only if both indicators designate service provision. Second, we observe data on in-house benefits provisions from the Woodrow Wilson Rehabilitation Center (WWRC), a state agency that provides comprehensive, individualized services with an employment objective. The WWRC receives an annual block grant from DARS which it administers autonomously. When appropriate, DARS refers individuals to WWRC for rehabilitative services. The WWRC provided us with service information for this type of in-house benefit. Because there may be some classification errors between in-house services and similar benefits, we identify them simply as "non-purchased services." These two sources of information cover all non-purchased service expenses except for in-house counselor services.

The first column of Table 2 shows the proportion of the sample receiving purchased services while the last column shows the prevalence in the sample for those receiving non-purchased services only.

These measures of purchased and non-purchased service receipt are used in the service receipt equation (equation (1)), the labor market equations (equations (2) and (3)), and the DI/SSI equation (equation (4)). However, if the only source of service receipt is in-house and/or similar benefits, then the β_j coefficients in equation (1) are multiplied by a service-choice "in-house service/similar benefits" parameter (to be estimated), and the $\left(\alpha_{jk}^z, \alpha_{jk}^w, \alpha_{jk}^r\right)$ coefficients in equations (2), (3), and (4) are multiplied by an outcomes "in-house

Table 2: Proportion Receiving DRS Purchased and Non-Purchased Service by Type

Service Type	Purchased Services			Non-Purchased Services Only
	Total	w/ Positive Expenditures	w/ Zero Expenditures	
Diagnosis & Evaluation	0.547	0.477	0.070	0.060
Training	0.372	0.362	0.010	0.100
Education	0.132	0.131	0.001	0.156
Restoration	0.319	0.318	0.001	0.108
Maintenance	0.301	0.300	0.001	0.123
Other Service	0.234	0.232	0.002	0.140

service/similar benefits" parameter (to be estimated). This allows both the service choice decisions, labor market, and DI/SSI outcomes to depend upon the source of the service. The first column of Table 2 shows the proportion of the sample receiving purchased services while the last column shows the prevalence in the sample for those receiving non-purchased services only.[9]

In addition to classifying services by provider (purchased and non-purchased), we also classify them by type. There are 76 separate services provided by DARS, other state agencies, and 1252 vendors.[10] Following Dean et al. (2002), we aggregate these services into the six service types, $DTERMO$, listed in Table 2.[11] As discussed above, *diagnosis & evaluation*[12] are provided at intake in assessing eligibility and developing an IPE and possibly later in the form of job counseling and placement services. *Training* includes vocationally-oriented expenditures including those for on-the-job training, job coach training, work adjustment, and supported employment. *Education* includes tuition and fees for a GED (graduate equivalency degree) program, a vocational or business school, a community college, or a university. *Restoration* covers a wide variety of medical expenditures including dental services, hearing/speech services, eye-

---

[9] If a particular service type is in both sources, we report the individual as receiving the service only once.

[10] Of the 1252 vendors, 73 are employment service organizations which receive roughly half of total purchased-service dollars, usually in the form of job coach services or supported employment.

[11] Although purchased services and in-house services provided by WWRC map uniquely into $DTERMO$, 4 of the 22 categories used for the RSA-911 do not. For example, the RSA category *diagnostic & treatment* ($D\&T$) includes both the *diagnosis & evaluation* category as well as the *restoration* category. Using $D\&T$ as an example, 6 of the 75 DARS purchased service categories map into *diagnosis & evaluation*, and 14 map into *restoration*. For the individuals flagged by RSA codes as having received $D\&T$, we count the number of sample individuals who received a service in one or more of the 6 *diagnosis & evaluation* purchased service codes (D) and the number of sample individuals in one or more of the 14 *restoration* codes (R). We then assign a probability that an individual designated in the RSA-911 file as receiving $D\&T$ receives *diagnosis & evaluation* as $0.56 = D/(D + R)$ and restoration as $0.44 = R/(D + R)$.

[12] We put variable names in a different font to avoid confusion.

glasses and contact lenses, drug and alcohol treatments, psychological services, surgical procedures, hospitalization, prosthetic devices, and other assistive devices. *Maintenance* includes cash payments to facilitate everyday living and covers such items as transportation, clothing, motor vehicle and/or home modifications, and services to family members. *Other services* consists of payments outside of the previous categories such as for tools and equipment.

Diagnostic and evaluation services are purchased in 54.7% of the base cases. Purchased services are provided in less than 40% of the cases for every other service type. This should be qualified by noting that 16% of applicants are not accepted into the program, and another 30% drop out after acceptance but before receiving substantive services. Of the remaining applicants, 80% are provided a purchased service other than for *diagnosis & evaluation*. The second column of Table 2 provides information on the proportion of individuals who receive non-purchased services. With the exception of *diagnosis & evaluation*, the frequency of the receipt of non-purchased services is very small with the exception of *diagnosis & evaluation*.

A high proportion of clients receive multiple services during the same service spell. For example, while the most common service combination in the initial service spell is *diagnosis & evaluation* with no other service, the next most common is *diagnosis & evaluation* along with *restoration*, and *diagnosis & evaluation* along with *training* is the fourth most common. Given the frequency with which clients receive multiple services, it is critical for us to allow for the possibility of receipt of multiple services. Thus, the structure of the service choice in equation (1) is multivariate discrete choice rather than polychotomous discrete choice.

Throughout much of the analysis, we measure labor market outcomes relative to the initial service period, defined as the first quarter in which purchased services are provided.[13] While this is a simple and appealing way to define the date of service receipt, there are two potential shortcomings of this measure: first, the initial service quarter may differ from the application quarter, and second, some clients receive services over multiple quarters. Figure 1 provides information about the importance of these issues, with the curve labeled "Case Open vs Assumed Service Date" revealing the density of how long it takes

---

[13]We construct the initial service period as lasting one quarter and with the quarter of service t_s defined in the following way: Define a_s as the starting quarter of the initial case and a_e as the ending quarter of the initial case. Let $s_{it} = 1$ iff i received purchased service in quarter t for $a_s \leq t \leq a_e$. Then

$$t_s = \begin{cases} \min_{a_s \leq t \leq a_e} t : s_{it} = 1 & \text{if } \sum_{a_s \leq t \leq a_e} s_{it} > 0 \\ a_s & \text{if } \sum_{a_s \leq t \leq a_e} s_{it} = 0 \end{cases} ;$$

i.e., the quarter of service is the first quarter during the initial spell when purchased service is received or the first quarter of the initial spell when no service is received. In Figure 1, the curve labeled "Case Open vs Assumed Service Date" is the density of $\Delta_s = t_s - a_s$ and the curve labeled "Assumed Service Date vs Last Service Receipt" is the density of

$$\Delta_e = \left[\max_{a_s \leq t \leq a_e} t : s_{it} = 1 \right] - t_s.$$

10

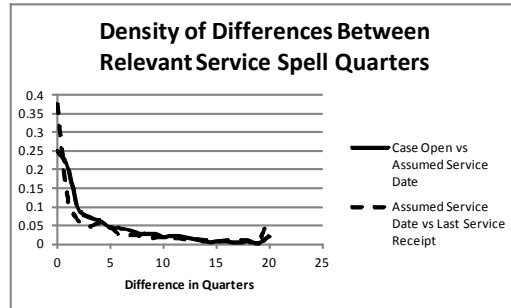**Density of Differences Between Relevant Service Spell Quarters**

Figure 1: Density of Differences Between Relevant Service Spell Quarters

(in quarters) to start receiving service after the application quarter, and the curve labeled "Assumed Service Date vs Last Service Receipt" displaying the density of the length of service receipt.[14] The first issue associated with the difference between the application and service dates is that one might want to treat labor market outcomes differently before and after application quarter (e.g., the Ashenfelter dip). Instead, we focus on a one-quarter pre-service dip in our specification of the model (see Section 2).The figure shows that 44% start receiving services in the application quarter and 83% start within 2 quarters. Meanwhile, 3% of DARS clients receive initial services 12 or more quarters after the application date. Thus, this issue may not matter that much given the concentration near zero. The second issue associated with the length of spells is that there may be a significant difference in labor market outcomes while service is being received and after it is finished. In our specification of the model, we distinguish between outcomes 8 or fewer quarters after service and 9 or more quarters after service. Figure 1 shows that 56.1% receive services for 3 quarters or less and only 19.1% of applicants are still receiving service after 8 quarters. Thus, for the most part, one can interpret the results for 9 or more quarters as being post-service receipt.[15]

---

[14]The up-tick at 20 quarters occurs because of censoring imposed by us at 20 quarters for this figure.

[15]One alternative way to define post-service outcomes would be to use the closing date of the service spell as the end of service. This is the case for most of the literature (e.g., Dean and Dolan, 1991). The problem with this approach is that counselors do not close cases necessarily when service provision ends. Another way is to model the transition associated with the end of purchased service receipt. We think this is an important long-term research goal but beyond the goals of this paper. Alternatively, one could just use the end of service receipt as the quarter defining the beginning of relevant labor market outcomes; we were somewhat concerned with endogeneity issues associated with the length of service receipt and later labor market outcomes. While our approach has issues associated with it as well, its simplicity makes it a good place to start exploration of the data.

3.1.3 DARS Data for Explanatory Variables

Table 3 provides the sample moments for the explanatory variables coming from the DARS data to be used in the analysis. While many of the variables are standard for this type of analysis, some are unusual and included because of the nature of the people being considered. *Special education* is a dummy variable equal to 1 for those observations where the respondent received some type of special education; 2.5% of the respondents received such education. Education information is missing for 10.3% of the sample. Rather than exclude such observations, instead we included a dummy variable for when education information was missing.

There are a number of indicators of physical and mental disabilities in the DARS data. We use four dummy variables, each equal to one if the individual's primary or secondary disability at intake in the base SFY 2000 case was diagnosed as a musculoskeletal impairment, a learning disability, a mental illness, and a substance abuse problem.[16] The meaning of mental illness as an explanatory variable is that, at the time of application to DARS, the individual's primary or secondary diagnosis was mental illness.[17] An individual's counselor also assesses the significance of the disability. Three levels are identified: *not significant* (used as the base level), *significant*, and *most significant*. We also constructed a dummy for serious mental illness (SMI) based on detailed diagnostic codes.[18]

While some variables such as *married* and *# dependents* may be endogenous, we follow the literature (e.g., Keith, Regier, and Rae, 1991; Ettner, Frank, and Kessler, 1997) and include them anyway as significant indicators of inclusion in society and responsibility. We include a dummy for receipt of government financial assistance even though it may be endogenous. However, for this population, one can work without losing one's government assistance or having it reduced up to relatively high earnings thresholds (see Figure 6). Finally, we include two transportation variables: *transportation available* and *has driver's license*. Raphael and Rice (2002) worries about the endogeneity of these variables and finds that controlling for endogeneity with some reasonable instruments has little effect on the estimated effect of transportation on employment but makes its effect on wages disappear.

To identify the impact of services on labor market outcomes and DI/SSI receipt, we exploit two instrumental variables that are correlated with the treat-

---

[16] The existence of visual, hearing/speech, internal disabilities, and other miscellaneous disabilities and cognitive impairments were available in the data but not common enough or not varying enough with dependent variables to measure precise effects. So they were not used in the analysis.

[17] Those in the sample with this explanatory variable equal to zero have a primary or secondary diagnosis of mental illness at some other point of re-entry into the DARS system in the future.

[18] Sometimes, SMI is defined in terms of the loss of functionality caused by mental health problems (U.S. Dept. HHS, 1999). Our data do not provide us with enough information to construct such a measure. Instead, we just take the two mental health diagnoses, schizophrenia and psychosis, among those we observe and define them, for this paper, as having SMI. It is essentially a parsimonious way to aggregate more specific mental health diagnoses.

Table 3: Moments of Explanatory Variables					
Socio-Demographic Variables			Disability Variables		
Variable	Mean	Std Dev	Variable	Mean	Std Dev
Male	0.404	0.491	Type		
White	0.710	0.454	Musculoskeletal Disability	0.170	0.376
Education	10.718	4.931	Learning Disability	0.046	0.209
Special Education	0.025	0.156	Mental Illness	0.950	0.218
Education Missing	0.103	0.304	Substance Abuse Problem	0.151	0.358
Age (Quarters/100)	1.427	0.407			
Married	0.178	0.383	Extent		
# Dependents	0.804	1.171	Significant	0.619	0.486
Transportation Available	0.741	0.438	Most Significant	0.275	0.446
Has Driving License	0.678	0.467	SMI	0.236	0.425
Receives Govt Assistance	0.191	0.300			

ment assignment but not included in the labor market equations (2) and (3) or the DI/SSI receipt equation (4). These instruments are the proportion of other clients in our cohort for the individual's counselor receiving a particular service and the proportion of other clients in our cohort for the individual's field office receiving a particular service. These variables are transformed as is described in Appendix 8.2.

The properties of these instruments depend upon the distribution of client size in our sample across counselors and field offices[19] and the distribution of the proportion of clients receiving each service. Figures 2 and 3 provide some information about these distributions. In Figure 2, we see that there is significant variation in the size of counselor caseloads and field office caseloads. For example, 43% of counselors have caseloads from our cohort of 5 or less, and 7.3% have caseloads of 20 or more. Analogously, 36.7% of field offices have caseloads from our cohort of 10 or less, and 20.5% have caseloads of 50 or more.

Figure 3 shows the empirical distribution of proportion of clients for each field office receiving each service. For example, for *diagnosis & evaluation*, 10.4% provide the service to 18.2% of their clients or less, and 4.2% provide it for all of their clients. Figure 3 shows that *diagnosis & evaluation* is the most commonly provided service, followed by *training*, then *restoration* and *maintenance*, then *other services*, and then *education*. In fact, except for *restoration* and *maintenance* and some choices at very low levels of provision, each curve stochastically dominates the ones behind it across offices. The distributions for counselors have similar properties.

There is strong evidence of important variation in behavior across counselors and across field offices.[20] We reject the null hypothesis that the joint density of services within offices does not vary across offices using a likelihood ratio test. The test statistic is 407.44 (with 245 df and normalized value of 7.33). We also can test the null hypothesis that each office provides each service in the same proportion, one at a time, using a likelihood ratio test. The test

---

[19] For example, if most counselors and/or field offices had only one client, then this methodology would not be useful.

[20] While there is significant positive correlation across counselor and office effects, there is enough independent variation between them to accurately estimate their effects on service provision.
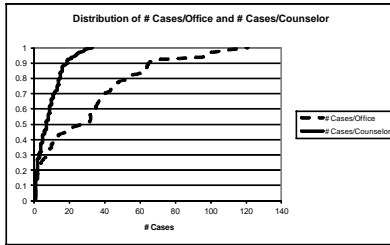
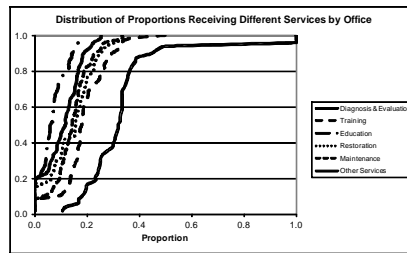Figure 2: Distribution of # Cases/Office and # Cases/Counselor

Figure 3: Distribution of Proportions Receiving Different Services by Office

statistic is 575.39 (with 294 df and a normalized value of 11.60). For counselors, the analogous test statistics are 970.60 (with 785 df and a normalized value of 4.68) and 3836.94 (with 942 df and a normalized value of 66.70). The fact that there is significant variation in the provision of services across offices and counselors make our instrument viable. Whether these instruments satisfy other identification restrictions is evaluated in Section 4.2.

3.2 VEC Data

One of the unique and valuable features of this analysis is that we have information from an administrative data source about individual quarterly earnings prior to, during, and after service receipt. Earlier economic analyses of VR efficacy (Conley, 1969; Bellante, 1972; Worrall, 1978; Nowak, 1983) relied almost exclusively on the RSA-911 Case Service Report of nationwide closures from the VR program. At the time, the RSA-911 form provides self-reported weekly earnings only at two points: 1) at the time of referral to the VR program and 2) following two months of employment. The latter figure is available only for that portion of VR cases closed "with an employment outcome." More recent analyses, published almost entirely in the rehabilitation literature (e.g., Cimera, 2010), utilize the same RSA-911 earnings measure, albeit now collected after three months of employment. In contrast, this study uses data gleaned from quarterly employment records provided by employers to the Virginia Employment Commission (VEC) for purposes of determining eligibility for unemployment insurance benefits.

14

The DARS provided the VEC with identifiers from the universe of 10323 applicants for DARS services in SFY 2000. The VEC returned to DARS a longitudinal file containing employment data for 9041 individuals having at least one quarter of "covered" employment during the 47-quarter period spanning July 1995 through March 2009, a "hit rate" of 88%. The remaining 12% in this cohort were either a) unemployed or out of the labor force for this entire interval or b) employed in jobs that are not covered by the VEC (e.g., were self-employed or worked out of state, for federal employers, for very small-sized firms, or at contingent-type jobs that do not provide benefits).

We explored the coverage issue through an arrangement with the Social Security Administration (SSA) whereby they matched VEC earnings (aggregated to a calendar year) to calendar-year SSA earnings for all SFY 2000 applicants.[21] Table 4 summarizes these results for the 9913 individuals with an identification match. For the two calendar years following SFY 2000 (the fiscal year of application), the SSA and VEC agreed on employment status for 87% of individuals. VEC records missed employment covered by SSA for 12% of the individuals in both 2001 and 2002. For those individuals where both SSA and VEC report earnings, VEC earnings levels fall short of SSA levels by 5.6% in 2001 and 6.2% in 2002.[22] Although formally accounting for these coverage errors is beyond the scope of this paper, the results in Table 4 suggests that any resulting biases should be minimal for the earnings equations but may be more important for the employment regressions. Unfortunately, our agreement with the SSA did not allow us to assess whether these errors varied by VR service receipt. If the errors are exogenous, the resulting estimates will be consistent. Otherwise, the extent of the bias in non-linear models is difficult to assess, especially when the errors are systematic.

Employers report aggregate earnings in a given quarter to the VEC. Recall that equations (2) and (3) model employment and earnings impacts in four separate periods offset from the date of first service. Because the date of first service can fall anywhere within a quarter, that quarter is excluded from the analysis other than for use as a period of demarcation separating pre-service from post-service periods. Depending upon the date of first service, this alignment procedure results in 16 to 19 quarters of pre-service earnings periods and 28 to 31 quarters post-service quarters for individuals in this cohort.

In our analysis, we try to explain two labor market outcome variables: *em-*

---

[21] This analysis was not limited to applicants with mental illness diagnoses.

[22] Data from the National Health Interview Survey 2004 Adult Sample (NHIS) show that, for the United States as a whole, people with mental illness have probabilities of working for the federal government and being self-employed of 2.7% and 7.8%, respectively; corresponding numbers for those without mental illness are 3.0% and 8.4%, respectively. However, because of its proximity to Washington, DC and its large number of military facilities, Virginia has an unusually high proportion of federal workers. Using data from the Bureau of Economic Analysis (2010b), the proportion of employed individuals in Virginia working for the federal government (including the military) in 2000 was 7.6%, while the NHIS data implies that it was 3.3% for the United States in 2004. If we conclude that $7.8\% + (7.6/3.3) * 2.7\% = 14.2\%$ of Virginians with mental illness either work for the federal government or are self-employed, this accounts for all of the discrepancy between SSA earnings and VEC earnings in the 2nd row of Table 4.

Table 4: Comparison between SSA and VEC
Employment Records

	2001	2002
Neither SSA nor VEC show earnings	31%	35%
SSA shows earnings, VEC does not	12%	12%
VEC shows earnings, SSA does not	1%	1%
Both SSA & VEC show earnings	57%	52%
Mean SSA Earnings	$9,117	$9,859
Mean SSA - VEC Difference	$510	$616

Table 5: Moments of Employment and Earnings Variables

	Before Initial Service Quarter			After Initial Service Quarter		
Variable	# Obs	Mean	Std Dev	# Obs	Mean	Std Dev
Employment	31427	0.35	0.477	58763	0.275	0.446
Log Quarterly Earnings	11003	7.082	1.492	16145	7.519	1.398

ployment and *log quarterly earnings*. *Employment* is a binary measure of working in a particular quarter in the labor market and is modeled in equation (2). We also measure *log quarterly earnings* in equation (3). While it would be valuable to be able to decompose quarterly earnings into wage level and hours, this is not possible in the VEC data. Table 5 provides information on sample sizes and on the moments of employment data and earnings data disaggregated between quarters before and after initial service provision. The sample sizes are quite large and allow us to estimate labor market outcome effects with high precision. One can see that employment rates decline after service provision and quarterly earnings increase (conditional on working). However, as is shown in Section 5, these aggregate facts hide what is really happening and how it depends on service receipt.

Figures 4 and 5 display quarterly employment rates and earnings (conditional on employment), respectively, for SFY 2000 applicants who receive substantial VR services and those that do not receive substantial services. We refer to these two groups as the treated and untreated, respectively. In these figures, quarters are measured relative to application date (not the initial service date) so that quarter 0 is the quarter of application, quarter -4 is one year prior to application, and quarter 4 is one year post-application.

Perhaps the most striking finding is seen in Figure 4 which shows that, prior to the application quarter, the employment rates of the treated and untreated are nearly identical, with a modest Ashenfelter dip in the pre-application quarter, but, just after the application quarter, the treated experience a pronounced increase in employment rates. For example, one year prior to the application quarter, the employment rates are 0.42 for both the untreated and treated, while, one year after the application, the analogous employment rates are 0.35 for the untreated and 0.46 for the treated. About one year after the application, the employment rates for both the treated and untreated start to decline, but a
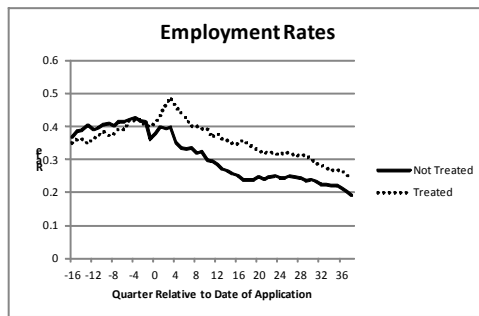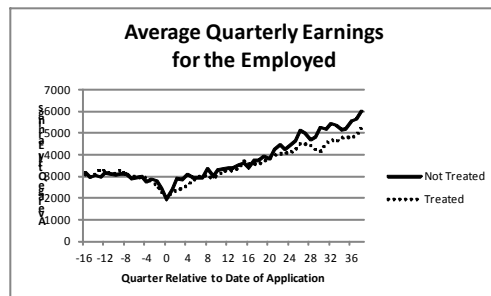
Figure 4: Employment Rates

Figure 5: Average Quarterly Earnings for the Employed

17

gap continues between the two groups. After nine years, the employment rates of around 0.20 are notably less than the rates in SFY 2000.

While there is notable association between DARS services receipt and employment, there is no such relationship with earnings. Figure 5 shows that quarterly earnings among the employed are almost identical for the treated and the untreated throughout. Thus, the data reveal that VR treatment services are associated with a sharp, substantial, and sustained increase in employment but no discernible change in quarterly earnings among the employed. While these results may suggest that VR programs effectively increase employment, we caution against drawing this type of causal conclusion from this evidence alone. The observed post-application increase in employment rates for treated clients may be due to VR services, but it may also reflect endogenous factors. This selection problem will be addressed using the structural model developed in Section 2.

Figures 4 and 5 also shed some light on the appropriate assumption about the length of the Ashenfelter dip. Depending on the program being evaluated, the pre-program dip in employment and earnings has been generally found to start between one quarter and one year prior to participation in the program (Heckman et al., 1999; Mueser et al., 2007). For our sample, Figures 4 and 5 reveal a dip in earnings in the first quarter prior to the initial service receipt. Thus, we account for the Ashenfelter dip using a one quarter pre-service indicator in employment and earnings equations.

3.3   SSA Data

A Memorandum of Agreement (MOA) between the Social Security Administration (SSA) and the Virginia Department of Aging and Rehabilitative Services (DARS) allowed us to obtain monthly SSI (Supplemental Security Income) and DI (Disability Insurance) payments for individuals in our cohort over the period of our analysis. We aggregate SSI and SSDI benefit recipiency into a single binary indicator of receipt of disability benefits from either source. We do this because the criterion for determining which of the two one receives is about prior accumulation of social security benefits, and, for this paper, that is a second-order issue. In our sample, we observe receipt of disability benefits in 38.4% of the person/quarters. The correlation of disability benefit receipt and work activity (measured as a binary variable) is -0.293, and the person-specific correlation is -0.189.

Figure 6 shows the distribution of the difference between quarterly earnings and the threshhold where one would start losing benefits (SGA). It is clear that very few in the sample treat the SGA as a binding constraint; i.e., there is no "parking" right below the SGA. This suggests that government receipt of benefits might not affect earnings in the way economists think, at least for this population at this period of time.
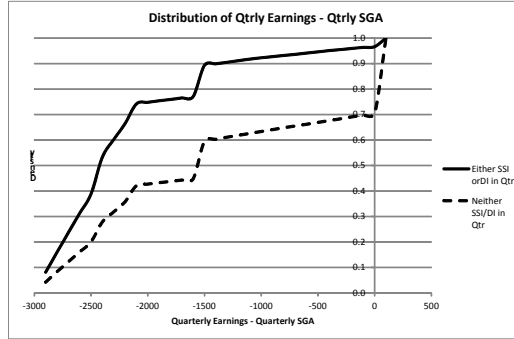
18

Figure 6: Distribution of Quarterly Earnings - Quarterly SGA

3.4 BEA Data

Labor market outcomes may be influenced by local labor market conditions. Though there are no measures of local labor market conditions in either the DARS data or the VEC data, the DARS data contain geographic identifiers so that we can match each DARS client with their county of residence. The Bureau of Economic Analysis (BEA) provides information on population size and number of people employed, disaggregated by age and county (BEA, 2010a). We construct measures of log employment rates using county level data. Details are included in Appendix 8.3.

4 Econometric Methodology

4.1 Likelihood Function

The parameters of the model are $\theta = (\theta_y, \theta_z, \theta_w, \theta_r, \Omega_\zeta)$ where

$$
\begin{aligned}
\theta_y &= \left(\beta_j, \lambda_{j1}^y, \lambda_{j2}^y \right)_{j=1}^J, \\
\theta_z &= \left(\gamma, \lambda_1^z, \lambda_2^z, \rho_\eta^z, \left[\alpha_{jk}^z \right]_{j=1}^J \right), \\
\theta_w &= \left(\delta, \lambda_1^w, \lambda_2^w, \rho_\eta^w, \sigma_w^2, \left[\alpha_{jk}^w \right]_{j=1}^J \right), \text{ and} \\
\theta_r &= \left(\psi, \lambda_1^r, \lambda_2^r, \rho_\eta^r, \left[\alpha_{jk}^r \right]_{j=1}^J \right).
\end{aligned}
$$

We estimate the parameters of the model using maximum simulated likelihood (MSL). The likelihood contribution for observation i is

$$
L_i = \int L_i(u_i) \, dG(u_i \mid \Omega) \tag{6}
$$

where

$$
L_i(u_i) = L_i^y(u_i^y) \prod_{t=1}^T L_{it}^{zwr}(u_{it}^z, u_{it}^w, u_{it}^r),
$$

19

$$L_i^y \left(u_i^y \right) = \prod_{j=1}^{J} \frac{\exp \left\{ X_i^y \beta_j + u_{ij}^y \right\}}{1 + \exp \left\{ X_i^y \beta_j + u_{ij}^y \right\}}, \tag{7}$$

$$L_{it}^{zwr} \left(u_{it}^z, u_{it}^w, u_{it}^r \right) = \left[L_{it}^0 \left(u_{it}^z, u_{it}^w \right) \right]^{1-z_{it}} \left[L_{it}^1 \left(u_{it}^z, u_{it}^w \right) \right]^{z_{it}} L_{it}^2 \left(u_{it}^r \right), \tag{8}$$

$$L_{it}^0 \left(u_{it}^z, u_{it}^w \right) = 1 - \Phi \left(X_{it}^z \gamma + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^z y_{ij} + u_{ijt}^z \right), \tag{9}$$

$$L_{it}^1 \left(u_{it}^z, u_{it}^w \right) = \frac{1}{\sigma_w} \phi \left(\frac{w_{it} - X_{it}^w \delta - \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^w y_{ij} - u_{it}^w}{\sigma_w} \right) \cdot \tag{10}$$
$$\Phi \left(X_{it}^z \gamma + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^z y_{ij} + u_{ijt}^z \right),$$

$$L_{it}^2 \left(u_{it}^r \right) = \Phi \left(X_{it}^r \psi + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^r y_{ij} + u_{ijt}^r \right)^{r_{it}} \cdot \tag{11}$$
$$\left[1 - \Phi \left(X_{it}^r \psi + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^r y_{ij} + u_{ijt}^r \right) \right]^{1-r_{it}},$$

and $G\left(u_i \mid \Omega \right)$ is the joint normal density with covariance matrix Ω described in Appendix 8.1. While, in general, it is difficult to evaluate the multivariate integral in equation (6), it is straightforward to simulate the integral using well-known methods described in Stern (1997). The functional form of the conditional likelihood contribution associated with observed program choices, $L_i^y \left(u_i^y \right)$ in equation (7), follows from the assumption in equation (1) that the idiosyncratic errors are iid logit. The functional form of the conditional likelihood contribution for labor market outcomes and DI/SSI receipt, $L_{it}^{zwr} \left(u_{it}^z, u_{it}^w, u_{it}^r \right)$ in equations (8), (9), (10), and (11) follow from the normality assumption for $(v_{it}^z, v_{it}^w, v_{it}^r)$ and the trivariate normality assumption for $(\zeta_{it}^z, \zeta_{it}^w, \zeta_{it}^r)$ in equation (5). The log likelihood function is

$$L = \sum_{i=1}^{n} \log L_i.$$

In theory, the parameter estimates are consistent only as the number of independent draws used to simulate the likelihood contributions goes off to infinity. However, Börsch-Supan and Hajivassiliou (1992) shows that MSL estimates perform well for small and moderate numbers of draws as long as good simulation methods are used,[23] and Geweke (1988) shows that the simulation error occur-

---

[23] We simulate all errors except for η and ε with antithetic acceleration (Geweke, 1988) and then compute likelihood contributions condition on the simulated errors. This is similar to simulation methods described in Stern (1992) and McFadden and Train (2000).

ring in simulation-based estimators is of order $(1/n)$ when antithetic acceleration is used.

4.2 Identification

There are two relevant notions of identification in this model. First, there is the general question of identification of model parameters in any nonlinear model. Second, service receipt, labor market outcome variables, and DI/SSI receipt are likely to be endogenous. With respect to the first issue, covariation in the data between dependent variables and explanatory variables identifies many of the model parameters. For example, covariation between *male* and participation in training identifies the β_j coefficient in equation (1) associated with the *male* for $j = training$. Similarly, the covariation between *white* and employment status identifies the γ coefficient in equation (2) associated with *white*, and the covariation between *white* and log quarterly earnings identifies the δ coefficient in equation (3) associated with *white*. Similar sample covariances identify parameters for DI/SSI recipiency in equation (4). Second moment parameters such as Ω_ζ in equation (5) are identified by corresponding second sample moments.

Two approaches are used to address the second identification problem. First, as in a difference-in-difference design, we control for pre-treatment labor market differences and DI/SSI receipt between those who do and do not receive services. If the differences in unobserved factors that confound inference in equations (2), (3) and (4), u_{it}, are fixed over time, then controls for the observed pre-treatment labor market differences and DI/SSI receipt address the endogenous selection problem (see Meyer, 1995; Heckman, LaLonde, and Smith, 1999, Section 4).

Second, we include two instruments in equation (1) that are excluded from equations (2), (3) and (4). As described in Sections 3.1.3 and Appendix 8.2, our choice of instruments for service j is the propensity of an individual's counselor to assign other clients to service j and the propensity of an individual's field office to assign other clients to service j.[24] Excluding these instruments from the labor market and DI/SSI equations seems sensible, and, as illustrated in Section 3.1.3, they are strongly correlated with service receipt. However, for these variables to be valid instruments it must also be the case that they are exogenous. While one can never be certain this holds, there are good reasons to think it is a reasonable assumption especially given that we include in the analysis the client's observed limitations, county-level employment rates, and pre-service labor market outcomes. Most notably, DARS clients have limited ability to select their field office or counselor; the field office is determined by the residential location of the client, and, conditional on observed limitations, counselors are randomly assigned.[25] So, unless clients relocate to take advantage of the practices of particular field offices, the assignment to offices and counselors

---

[24] Doyle (2007), Arrighi et al. (2010), Clapp et al. (2010), Maestas, Mullen, and Strand (2012), and Dean et al. (2013a, 2013b) and use a similar instrument in other applications.

[25] Counselors are assigned by office policy that does not involve client choice. For example, some field offices assign counselors to balance caseload across counselors, some have counselors who specialize in mental illness, and some assign counselors by client locale.

is effectively random conditional on the observed limitations of clients. A threat to the validity of these instruments may arise if variation in the availability of jobs where training (or other DARS services) is productive might jointly affect labor market outcomes and the average behavior of counselors and field offices. Including measures of local labor market conditions directly in equations (2) and (3) should ameliorate this problem. A final concern arises if there is significant unobserved variation in the ability of counselors to match clients with jobs, thus affecting both his/her decisions about what types service to offer clients and later success in the labor market. We assume that these types of effects are not important in our analysis.

Importantly, our approach for addressing the endogenous selection of services represents a substantial advance over the existing literature where the past research (often using RSA-911 data) generally relies on limited controls for pre-program earnings and assumes service participation is otherwise exogenous. Along with Aakvik, Heckman, and Vytlacil (2005) and Dean et al. (2013a, 2013b), this is the first study to identify the impact of VR services on labor market outcome using both a history of pre-program earnings and plausibly exogenous instrumental variables.

5 Estimation Results

5.1 Estimates of Impact of VR Services

We divide up the discussion of parameter estimates into separate components. We begin by examining the estimated effect of services on labor market outcomes. Table 6 presents the estimates and associated standard errors for the effect of services on employment, and Table 7 presents the analogous results for log quarterly earnings. For each labor market outcome, the effects are allowed to vary across the six different service types and across different time periods relative to the initial service quarter. Given our rich labor market data, we are able to estimate both short-run (the first two years) and long-run (more than two years) effects of services and account for pre-service outcomes in the quarter prior to services as well as two or more quarters prior to the initial service. As noted in Section 4.2, inclusion of pre-treatment periods is a way to account for the effect of endogenous selection into services. This method of controlling for selection, which is the central idea of the difference-in-difference design, is used extensively in the literature (e.g., Meyer, 1995; Heckman et al., 1999). The quarter immediately prior to initial service provision is separated out because this quarter seems likely to have a distinct impact on selection and because of the well-documented variation in labor market behaviors just prior to the application period – the Ashenfelter dip (Ashenfelter, 1978; Heckman et al., 1999).

The first two columns of Tables 6 and 7, which display estimates for the quarters prior to the initial service, reveal evidence that selection is endogenous. Nearly all of the coefficients associated with periods two or more quar-

22

ters prior to the initial service are substantial and statistically different than zero, the one exception being the coefficient on *restoration* in the *log quarterly earnings* equation. For *training* and *maintenance*, the estimates reveal that those individuals provided training services have lower pre-treatment employment probabilities and quarterly earnings. For *education* and *other services*, the estimates imply selection is positively associated with pre-service labor market outcomes – people with mental illness with higher pre-treatment employment rates and earnings are more likely to be assigned to these services. In general, the results for the quarter one period prior to services are qualitatively similar although in many cases are not statistically different than zero. Overall, these results suggest a complex and heterogeneous selection process where applicants are assigned to particular services based on underlying unobserved factors that are associated with pre-service labor market outcomes.

The last two columns of results display the estimated short- and long-run effects of services on labor market outcomes. These estimates should be interpreted relative to the coefficients associated with pre-service measures in the first two columns. For example, as seen in Table 6, prior to service provision, the *employment* propensity for clients provided training services is 0.361 less than for clients that do not receive these services.[26] In the two years after the start of service provision, it rises to 0.270,[27] and then, in the longer run, it declines to 0.180. Relative to individuals in the sample who received no training services, the long-term *employment* propensity (z_{it}^* in equation (2)) is 0.180 higher for those that received training. Thus, after accounting for selection into service, the long-term effect of *training* on those who were trained is $0.180 + 0.361 = 0.541$.

The employment and log-quarterly earnings effects of each service type across the four time periods can be observed more easily in Figures 7 and 8, respectively. Relative to employment propensities two or more quarters prior to service provision, we observe that *training* and *other services* increase *employment* propensity while *diagnosis* & *evaluation*, *education*, and *restoration* decrease *employment* propensity. *Maintenance* increases *employment* propensity in the short run but decreases it in the long run.

Figure 8 shows that, for earnings effects, *restoration, maintenance,* and *other services* increase conditional earnings, relative to earnings two or more quarters prior to service provision, in both the short and long run, and *diagnosis* & *evaluation, training,* and *education* decrease conditional earnings in the short run but increase them in the long run.[28] Dean and Dolan (1991) also find evidence of positive earnings effects in their earlier evaluation of VR services, although in some cases, especially for men, the results are not statistically significant. After using an instrumental variable to address the selection problem, Aakvik et al.

---

[26] Throughout this discussion, the effect of an explanatory variable on *employment* propensity means the partial derivative of the latent value associated with employment with respect to the explanatory variable.

[27] Recall that this 2-year period is one where those receiving services are in the program to various degrees and with varying durations.

[28] Almost all F-statistics testing for the joint significance of the short-term and long-term *log quarterly earnings* effects relative to the effect prior to program participation are statistically significant with p-values less than 0.0001.

Table 6: DRS Purchased Service Participation Effects on
Employment Propensity

Variable	Two or More Quarters Prior to Service Participation	Quarter Prior to Service Participation	First 2 Years After Service Participation	More than 2 Years After Service Participation
Diagnosis & Evaluation	0.280 **	0.091	0.052 **	-0.182 **
	(0.008)	(0.081)	(0.014)	(0.007)
Training	-0.361 **	-0.168 *	0.270 **	0.180 **
	(0.010)	(0.101)	(0.017)	(0.008)
Education	0.283 **	0.175	-0.016	0.170 **
	(0.014)	(0.134)	(0.025)	(0.010)
Restoration	0.275 **	0.461 **	0.258 **	0.148 **
	(0.010)	(0.092)	(.018)	(0.008)
Maintenance	-0.217 **	-0.222 **	-0.163 **	-0.291 **
	(0.011)	(0.113)	(0.019)	(0.009)
Other Services	0.156 **	-0.035	0.284 **	0.205 **
	(0.011)	(0.122)	(0.020)	(0.009)

Notes:

1. Standard errors are in parentheses.

2. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.
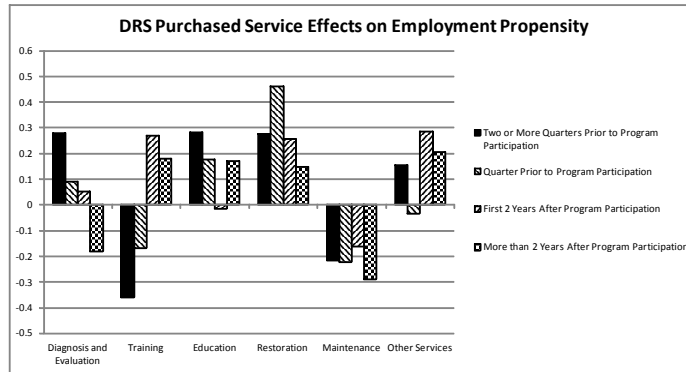

Figure 7: DRS Purchased Service Effects on Employment Propensity

Table 7: DRS Purchased Service Participation Effects on Log Quarterly Earnings

Variable	Two or More Quarters Prior to Service Participation	Quarter Prior to Service Participation	First 2 Years After Service Participation	More than 2 Years After Service Participation
Diagnosis & Evaluation	-0.017 **	-0.349 **	-0.102 **	0.015
	(0.013)	(0.107)	(0.025)	(0.011)
Training	-0.065 **	-0.086	-0.120 **	0.071 **
	(0.017)	(0.159)	(0.031)	(0.014)
Education	0.096 **	0.027	0.011	0.242 **
	(0.022)	(0.186)	(0.042)	(0.017)
Restoration	-0.028 *	-0.116	0.064 **	0.178 **
	(0.015)	(0.123)	(0.031)	(0.013)
Maintenance	-0.293 **	-0.027	-0.187 **	-0.076 **
	(0.018)	(0.169)	(0.033)	(0.013)
Other Services	0.070 **	-0.199	0.154 **	0.216 **
	(0.017)	(0.176)	(0.032)	(0.014)

Notes:
1. Estimates are effects on log quarterly earnings conditional on employment.
2. Standard errors are in parentheses.
3. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

(2005) find no evidence of employment effects of VR services in Norway.

Most previous evaluations of VR services focus on the impact of a single treatment indicator that is assumed to be conditionally exogenous. In this setting, the basic idea is to compare the differences in mean outcomes between treatment and control groups after conditioning on observed variables. For example, Figures 4 and 5 above, which display the unconditional mean employment and earnings outcomes respectively, reveal little pre-program differences, fairly substantial positive post-treatment employment associations, and almost no relationship between treatment and earnings. The structural model estimated in this paper extends this approach in several important ways: first, by conditioning on observed covariates; second, by accounting for six different types of service rather than a single treatment indicator; and finally, by using instrumental variables in a model with endogenous service provisions. The results from the structural model estimates presented in this section suggest a much more complex and nuanced story, with evidence of pre- and post-program labor market differences that vary across services, estimated employment effects that are positive for some services and negative for others, and estimated earnings effects that are consistently positive in the long run.

Because of the variation in effects over time and over labor market outcomes seen in Figures 7 and 8, it is difficult to infer the long-run benefits of each service. Accordingly, Figure 9 reports the mean present value for 10 years of earnings flows (measured in $1000) excluding service costs , a 95% confidence range,[29]

---

[29] The 95% confidence range provides information about the variation in benefits across individuals caused by the nonlinearity of the model and variation in other explanatory variables.
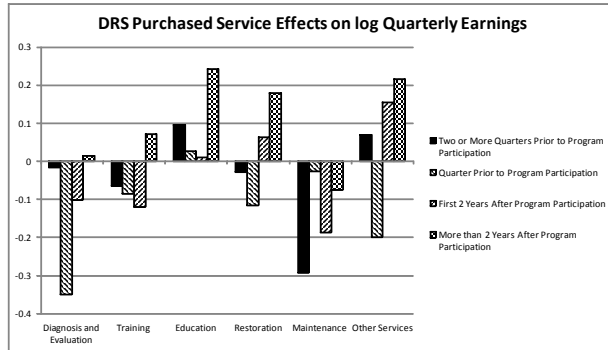
Figure 8: DRS Purchased Service Effects on log Quarterly Earnings

and the minimum and maximum present value of each service.[30]   Except for *diagnosis & evaluation*, all of the services have positive long-run benefits. On average, *training*, *restoration*, and *other services* have benefits on the order of $7200, $3750, and $4800 respectively, while *education* and *maintenance* have positive benefits of $1700 and $2100 respectively. It should be noted that, in Figures 7 and 8, *education* has a negative effect on both short- and long-run employment probabilities but a substantial long-run positive effect on quarterly earnings conditional on employment. Figure 9 shows that the long-run conditional earnings effects essentially offset the negative employment effects for present value calculations. One other notable feature of the discounted benefits calculations illustrated in Figure 9 is the high degree of variability across the caseload. The discounted benefits associated with training services, for example, range from $700 to nearly $22800. For the other service categories, there are notable fractions of the caseload that would receive negative benefits.

The negative long-run benefits for *diagnosis & evaluation* are somewhat difficult to understand.[31] This may reflect the fact that these services are largely provided in-house, yet our data do not fully reveal in-house services provided by counselors (see Section 3.1.2). So, while nearly every applicant receives some *diagnosis & evaluation* services, our data indicated that only 63% of applicants receive services – 49% purchased and 14% non-purchased (see Table 2). In addition to this measurement problem, purchased *diagnosis & evaluation* services may differ from other types of service in a number of ways, some of which imply that receipt of such services acts very much like a selection effect. In particular, purchased *diagnosis & evaluation* services tend to be provided for clients with

---

[30]We use a quarterly discount factor of 0.95. Because the distribution of benefits is highly skewed, the normal approximation is not appropriate. Instead, we report the 0.025 and 0.975 quantiles of the empirical distribution.

[31]We also found that *education* and *restoration* have negative outcomes in the short and long run for *employment* and positive long-run results for *conditional earnings*. For these services, service receipt appears to raise earnings and reservation wages. However, is not clear why reservation wages would rise faster than wages (which would be necessary for *employment* effects to be negative).
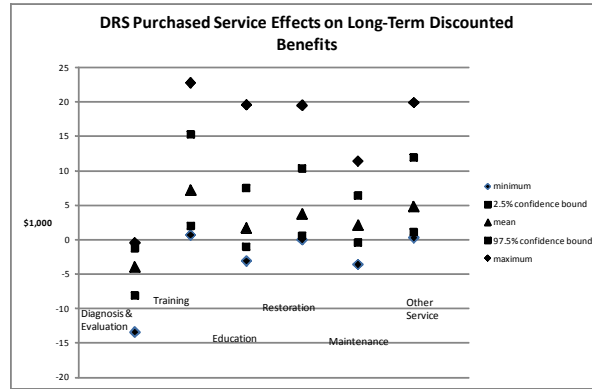
Figure 9: DRS Purchased Service Effects on Long-Term Discounted Benefits

especially difficult cases. In fact, 79% of those receiving purchased *diagnosis & evaluation* services are examined by specialists. Thus, unobserved heterogeneity in mental illness is an error in measuring an explanatory variable that is related to the provision of *diagnosis & evaluation* services and labor market outcomes. In this setting, under reasonable modeling restrictions, the instrumental variable estimate on *diagnosis & evaluation* will be negatively biased and the estimate on mental illness will be upward biased (see Appendix 8.4 for an illustration).

Another plausible explanation is that clients who receive purchased *diagnosis & evaluation* services are more likely to be diagnosed with problems that make it difficult for them to succeed in the labor market, and the DARS counselor influences them to move in a different, more rewarding direction. In such cases, while this would not lead to improved labor market outcomes, it might lead to the most productive outcome available. In fact, as discussed below, we find evidence that the purchased *diagnosis & evaluation* services increase the receipt of DI/SSI.

Finally, this may reflect a negative unobserved counselor specific effect; the least successful counselors may be the most likely to use purchased *diagnosis & evaluation* services and the least likely to succeed in helping their clients in the labor market.

Table 8 presents analogous results for the effect of services on DI/SSI recipiency, and Figure 10 provides the estimates graphically. The estimates show that every service increases the probability of DI/SSI receipt. For example, the long-term effect of *training* on the DI/SSI receipt propensity is estimated to be $0.423 - 0.124 = 0.299$, which implies that training services increase the probability of DI/SSI receipt on average by 0.05. Likewise, *diagnosis & evaluation* increase the probability of DI/SSI receipt by 0.18, *education* by 0.08, *restoration* by 0.06, *maintenance* by 0.03, and *other services* by 0.01. The effect for *diagnosis & evaluation* is the largest which lends more credence to the argument made earlier about the effect of *diagnosis & evaluation* on labor market outcomes. Much of the recent public policy debate about VR programs includes

Table 8: DRS Purchased Service Participation Effects on DI/SSI
Participation Propensity

Variable	Two or More Quarters Prior to Service Participation	Quarter Prior to Service Participation	First 2 Years After Service Participation	More than 2 Years After Service Participation
Diagnosis & Evaluation	-0.562 **	-0.139	0.244 **	0.667 **
	(0.014)	(0.151)	(0.019)	(0.008)
Training	0.124 **	0.574 **	0.592 **	0.423 **
	(0.015)	(0.189)	(0.027)	(0.009)
Education	-0.176 **	0.383	0.471 **	0.345 **
	(0.026)	(0.282)	(0.044)	(0.013)
Restoration	-0.691 **	-0.779 **	-0.699 **	-0.161 **
	(0.019)	(0.236)	(0.025)	(0.009)
Maintenance	-0.179 **	-0.009	0.193 **	0.054 **
	(0.016)	(0.202)	(0.028)	(0.010)
Other Services	-0.545 **	-0.318	-0.366 **	-0.391 **
	(0.018)	(0.230)	(0.031)	(0.010)

Notes:
1. Estimates are effects on DI/SSI Participation Propensity.
2. Standard errors are in parentheses.
3. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

some discussion on whether these programs might notably reduce the number of persons receiving disability benefits (e.g, see Hennessey and Muller, 1995; Autor and Duggan, 2010; Wilhelm and Robinson, 2010; Stapleton and Martin, 2012; Sosulski, Donnell, and Kim, 2012). Our estimates suggest this is unlikely to happen, at least for people with mental illness. Instead, the estimates in Tables 6 and 7 suggest that VR programs help some enter and succeed in the labor market, and the estimates in Table 8, along with the negative correlation between DI/SSI receipt and employment implied by the factor loadings for the first factor in Table 11 suggest that VR programs help those who are unlikely to succeed in the labor market apply for and receive disability benefits. From the point of view of reducing government expenditures, this is problematic. But, from the point of view of welfare, it is Pareto improving in that people with mental health problems are unambiguously better off and taxpayers are better off to the degree that they supported an efficient disability benefits program in the first place.

5.2 Estimates of the Impact of Covariates

Estimates of the effects of demographic characteristics on the propensity to use different services (y_{ijt}^* in equation (1)) are provided in Appendix 8.5. For the most part, the observed characteristics do not have statistically significant effects on service receipt, but there are some interesting exceptions. We find that clients with learning disabilities (0.680) and those receiving government assistance (0.491) are more likely to receive diagnosis & evaluation services. The probability of receiving *training* is higher for persons with government assistance (0.777) but lower for men (−0.315) and for those with musculoskeletal disabilities (−0.489) and/or substance abuse problems (−0.373). The receipt of *education* increases for those with more *education* (0.082) and for those with
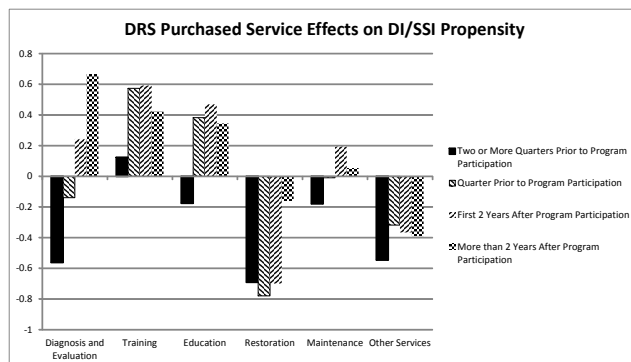
Figure 10: DRS Purchased Service Effects on DI/SSI Propensity

access to transportation and a driver's license (0.679). Interestingly, however, there is no statistically significant effect associated with having a serious mental illness or a significant disability.[32]

Table 9 presents estimates of counselor and office effects as defined in Appendix 8.2. There are two types of coefficient estimates reported in the table: a) the counselor and office effects and b) the missing counselor effects. The counselor and office effects should be interpreted as $\partial E y_{ij}^* / \partial e_i$ where y_{ij}^* is the latent variable associated with receipt of service j in equation (1) and e_i is the counselor or office effect defined in Appendix 8.2; note that these are restricted to be the same across different services. The missing counselor effects are the effect on y_{ij}^* when the relevant counselor does not have enough other clients to compute a set of counselor effects.[33] These counselor and office instrumental variables turn out to have large and statistically significant effects on service provision across clients. One should note that we are controlling for a pretty full set of demographic characteristics. So it is unlikely that these results reflect variation in the mix of clients across counselors and/or field offices.

Table 10 reports the effects of the demographic, socioeconomic, and disability-related characteristics on the three labor market outcomes of interest (z_{it}^* in equation (2), w_{it} in equation (3), and r_{it}^* in equation (4)). For labor market outcomes, almost all of the estimates are statistically significant. Many of the estimates are as expected including positive effects of being white on *employment* propensity (0.157) and *log quarterly earnings* (0.362) as well as positive effects of *education* on *employment* propensity (0.024) and *log quarterly earnings* (0.053). The two transportation variables also have positive impacts on

---

[32] The *education missing* variable is statistically significantly negative across almost all services. It turns out that almost all of the individuals with *education missing* were closed during the application process. Thus, in an important sense, causation for this variable runs the other way.

[33] We allow missing counselor effects to vary over services. However, we restrict missing office effects coefficients to be zero because there are not enough cases and those that exist are too highly correlated with missing counselor effects to estimate both with any precision.

Table 9: Counselor and Office Effects
on Service Receipt

	Estimate		Std Err
Counselor Effect	0.344	**	0.103
Office Effect	0.767	**	0.069
Missing Counselor Effects			
Diagnosis & Evaluation	-0.431		0.350
Training	0.105		0.421
Education	-0.753	*	0.433
Restoration	-0.486		0.362
Maintenance	-0.360		0.410
Other Services	-0.393		0.409

Notes:
1.Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.
2.Other than those reported, missing counselor and field office effects parameters were excluded because of multicollinearity problems.

both labor market outcomes. The local labor market employment rate increases employment probabilities but decreases conditional earnings, suggesting that it might have been useful to include a measure of local wage rates. Some of the demographic and socioeconomic parameter estimates are counterintuitive. In particular, having a serious mental illness (SMI) increases *employment* propensity (0.194), receipt of special education increases *log quarterly earnings* (0.259),[34] while being married decreases both *employment* propensity (-0.322) and *log quarterly earnings* (-0.138). The marriage effects can occur through income effects associated with having a spouse.

The diagnosis of a mental illness in the "base case" versus being initially diagnosed with mental illness in a subsequent application for VR services has a negative effect on *employment* propensity (-0.210) while increasing *log quarterly earnings* (0.399). Meanwhile, the disability severity-related variables have the expected signs, with negative effects of *significant* and *most significant* disabilities (relative to mild) on both labor market outcomes. Unlike its impact on service provision, the *SMI* estimates are explaining a significant amount of variation in labor market outcomes. *SMI*, by itself, increases employment (0.194) and increases *log quarterly earnings* (0.964). For males and whites, there are added interaction effects, all adversely affecting labor market outcomes. However, overall, the estimates with respect to *SMI* effects are hard to explain.[35] *Education* interacted with *SMI* has negative effects, and *age* interacted with *SMI* has mixed but statistically significant effects on outcomes. Baldwin (2005) estimates the effect of mood disorder, anxiety disorder, and adjustment disorder on employment probabilities and finds an average reduction in employment

---

[34] Special education programs have been found to improve schooling outcomes (Hanushek et al., 2002) and are associated with the use of supported employment services linked to higher earnings (Drake et al., 2009).

[35] The estimates imply that the average person in the sample has negative impacts of *SMI*. But the results are still problematic, for example, for black women with *SMI*.

Table 10: Labor Market and DI/SSI Effects

Variable	Employment		Log Quarterly Earnings		DI/SSI Receipt	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Constant	0.052 **	0.021	5.311 **	0.032	-3.338 **	0.026
Male	0.024 **	0.006	0.360 **	0.008	-0.036 **	0.007
White	0.157 **	0.007	0.362 **	0.009	-0.114 **	0.009
Education	0.024 **	0.001	0.053 **	0.001	-0.044 **	0.001
Special Education	-0.014	0.027	0.259 **	0.032	-0.019	0.024
Education Missing	0.423 **	0.013	0.626 **	0.020	-0.438 **	0.015
Age/100	-0.701 **	0.007	0.043 **	0.009	1.358 **	0.010
Married	-0.322 **	0.007	-0.138 **	0.009	0.484 **	0.009
# Dependents	0.061 **	0.002	0.064 **	0.003	-0.063 **	0.003
Transportation Available	0.122 **	0.007	0.089 **	0.009	-0.068 **	0.007
Has Driving License	0.213 **	0.007	0.333 **	0.010	-0.175 **	0.007
Receives Govt Assistance	-0.332 **	0.008	-0.237 **	0.012	1.753 **	0.009
Musculoskeletal Disability	0.097 **	0.007	0.130 **	0.009	0.036 **	0.009
Learning Disability	0.406 **	0.012	0.225 **	0.015	-0.166 **	0.015
Mental Illness	-0.210 **	0.012	0.399 **	0.018	-0.016	0.013
Substance Abuse	0.213 **	0.007	0.086 **	0.010	-0.366 **	0.009
Disability Significant	-0.027 **	0.009	-0.170 **	0.012	0.664 **	0.011
Disability Most Significant	-0.112 **	0.010	-0.280 **	0.013	1.147 **	0.012
SMI	0.194 **	0.028	0.964 **	0.038	1.083 **	0.032
Male * SMI	-0.094 **	0.012	-0.521 **	0.017	0.513 **	0.014
White * SMI	-0.535 **	0.012	-0.482 **	0.018	0.815 **	0.014
Education * SMI	-0.019 **	0.001	-0.047 **	0.002	0.044 **	0.001
Age/100 * SMI	0.236 **	0.015	-0.083 **	0.021	-1.249 **	0.016
Local Employment Rate	0.185 **	0.067	-0.185 **	0.068	0.050 **	0.011

Note: Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

probability on the order of 0.3 (see McKeithen and Stern, 2007, for calculations). Our estimates imply smaller effects, at least for significant mental health problems similar to those considered by Baldwin. A big part of the reason for this is probably that our sample consists only of people who have been identified as having a mental health problem while Baldwin (2005) uses the SIPP sample.

For DI/SSI receipt, almost all of the effects are statistically significant and with expected signs, also. For example, the probability a client takes-up DI/SSI is estimated to decrease for *white* (−0.114), *education* (−0.044), and *transportation available* (−0.068). Surprises are *married* (0.484), *learning disability* (−0.166), and *local employment rate* (0.050).

So far all of the discussion has concerned the effect of purchased services on labor market outcomes. In fact, DARS also provides some services in-house, and other services sometimes are paid for by other organizations, and, as discussed in Section 3.1.2, we have some information about those other services. Using this data, we allow the effects of covariates on the receipt of such services to be proportionate to their effect for service choice in equation (1) and their effect for employment propensity in equation (2) as reported in Table 6, for conditional log quarterly earnings in equation (3) as reported in Table 7, and for DI/SSI receipt in equation (4) as reported in Table 8. The estimated proportion for service choice propensity is 0.824** (0.280) which is not significantly different from 1.0. Thus, decisions about using non-purchased services are similar to those for purchased services. By contrast, the estimated proportion for employment propensity, conditional log quarterly earnings, and DI/SSI receipt propensity is 0.432** (0.033) which is significantly different from 1.0. Thus, the effect of non-purchased services on labor market outcomes and DI/SSI receipt is 43.2%

Table 11: Covariance Factor Loadings

	Factor 1		Factor 2	
Variable	Estimate	Std Err	Estimate	Std Err
Diagnosis & Evaluation	-0.055	0.065	-0.032	0.069
Training	0.006	0.085	-0.065	0.083
Education	-0.064	0.109	-0.152	0.110
Restoration	-0.039	0.075	-0.143 *	0.083
Maintenance	-0.056	0.086	0.197 **	0.088
Other Services	0.022	0.088	-0.257	
Employment	0.556 **	0.003	-0.002	0.003
Log Quarterly Earnings	0.436 **	0.004	-0.116 **	0.004
SSI/DI	-0.225 **	0.003	1.652 **	0.005

Notes:

1.Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level.

2.The identifying condition associated with the factor loadings is that the factor loadings for the six different services are orthogonal. We impose this condition by computing the factor loading for factor 2 on other services as a function of the other 11 relevant factor loadings. The factor loadings associated with labor market outcomes are not part of the orthogonality condition.

of that for purchased services.

5.3 Estimates of the Covariance Structure

Our model has a rich error covariance structure, as seen in equation (5). This allows for the possibility that unobservables associated with service provision are correlated with unobservables associated with labor market outcomes. The factor loadings for *Factor 1* in Table 11 demonstrate no statistically significant correlations between the errors associated with the provision of all service types and the error associated with *employment* propensity. However, the correlation between the errors for *employment* propensity and *log quarterly earnings* is positive (0.556 and 0.436) and the correlation between the errors for *employment* and *DI/SSI* receipt are negative (0.556 and −0.225). This suggests that there is some unobserved personal characteristic, maybe ability, that increases employment probabilities and conditional earnings but decreases DI/SSI receipt probabilities.

Meanwhile, the factor loadings for *Factor 2* imply negative correlation between the errors associated with *log quarterly earnings* and *DI/SSI* receipt (−0.116 and 1.652). This suggests another unobserved characteristic, perhaps some other component of ability, increasing *log quarterly earnings* and decreasing *DI/SSI* receipt but having no real impact on *employment* propensity. The estimates of the factor loadings for service provision imply that neither unobserved component has any meaningful effect on service provision except for *maintenance* which is positively correlated with *DI/SSI* receipt (0.197 and 1.652).

The estimates of the other elements of the error structure are reported in Ta-

Table 12: Other Covariance Terms

Variable	Estimate	Std Err	Variable	Estimate	Std Err
Var(ζ_1)	0.041 **	0.000	Var(ζ_2)	0.008 **	0.004
Cov(ζ_1,ζ_2)	0.018 **	0.005	Cov(ζ_2,ζ_3)	-0.018 **	0.005
Cov(ζ_1,ζ_3)	-0.040 **	0.000	Var(ζ_3)	0.041 **	0.000
		0.000	σ_w	1.281 **	0.003

Note: Double-starred items are statistically significant at the 5% level.

ble 12. All of the estimated covariance terms are relatively small and dominated by the factor structure terms in Table 11. The estimate of the log earnings error σ_w is quite large, implying that a standard deviation in quarterly earnings due to unobserved factors is on the order of \$8546. It is unclear how much of this variation is due to variation in wages and how much is due to variation in hours. Baldwin (2005) finds wage effects on the order of -0.2 (see McKeithen and Stern, 2007, for calculations) but does not estimate hours effects.

5.4 Specification Tests

We use standard goodness-of-fit tests to measure how well we are predicting service provision probabilities. For each service, we decompose the sample into 40 cells, each of length 0.025, stratified by the predicted probability of service receipt.[36] Then we construct the standard χ^2 test statistic. For service provision probabilities, we fail to reject the null that the model predictions equal observed probabilities at the 5% percent significant level.

We perform the same test for employment probabilities disaggregated into probabilities before and after service receipt.[37] The test statistics are $\chi^2_{31} = 288.29$ for employment probabilities before service receipt and $\chi^2_{32} = 109.87$ for employment probabilities after service receipt. Both of these are highly significant implying a poor fit. Figure 11 plots the deviations between predicted and sample employment probabilities for the two periods. Deviations between the 45° line and the other two *sample lines* at any particular predicted probability represent that part of employment probability that we are not predicting. The model does a good job predicting employment probabilities both periods up to a predicted probability of about 0.7, after which the fit worsens. Overall, while we are basically predicting employment probabilities reasonably well, there is some concern that, between 0.2 and 0.5, we are underestimating employment probabilities prior to VR service receipt and overestimating it after VR service receipt; this may have a large effect on our rate-of-return analysis later in Section 6.

Finally, for DI/SSI receipt, the test statistics are $\chi^2_{35} = 368.8$ for DI/SSI probabilities before service receipt and $\chi^2_{38} = 1218.1$ for DI/SSI probabilities

---

[36] For each test, some cells are empty and therefore not used.

[37] *Before service receipt* includes the quarter before receipt, and *after service receipt* includes both quarters in the first two years after receipt and the longer run.

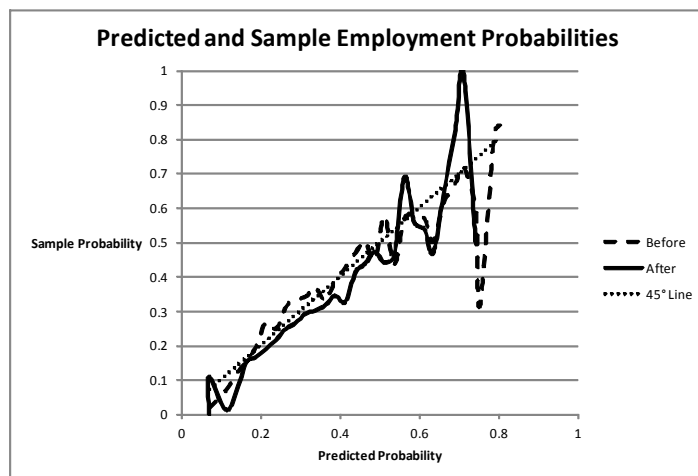**Predicted and Sample Employment Probabilities**

Figure 11: Predicted and Sample Employment Probabilities

after service receipt. However, the curves analogous to Figure 11 look extremely similar to those in Figure 11 (once smoothed);[38] they fit very well until predicted probabilities above 0.7 (where there many fewer observations).

Also, using a series of Lagrange Multiplier (LM) tests, we consider allowing for interactions among pairs of services in the labor market outcome equations. While the nonlinearity of the model creates some interactions, it may not be appropriate to rely strictly on the model structure. These LM tests, however, suggest that is not important to allow for service interactions. Similarly, we test for interactions between demographic characteristics (*male* and *white*) and services, but find no evidence of such interactions.

6 Rate of Return

The preceding analysis suggests that, except for purchased diagnosis & evaluation services, observed DARS services have long-run positive effects on labor market outcomes (see Figure 9). In this section, we examine the social welfare implications of VR services by comparing the estimated benefits and costs of the program. The primary monetary benefits and costs of VR services are estimated using our model and the DARS data on the costs of purchased services. There are, however, many factors for which we do not have direct evidence on the associated costs. In particular, the costs of non-purchased services are not observed in the DARS data file. For these items, we present more speculative evidence.

We simulate the private labor market benefits to DARS clients using the

---

[38] The graph for DI/SSI receipt is available in Appendix 8.6.

structural model estimates summarized in Section 5.[39] In particular, we compute the mean present discounted value of the provided services relative to the value of receiving no services using both a 5- and 10-year post-treatment observation period for those individuals who received some service and using an annual discount factor of 0.987.[40] The estimated mean discounted benefits are $1942 with a standard deviation of $3726 using the 5-year window and $4124 with a standard deviation of $7677 using a 10-year window.

A more accurate estimate of the mean discounted benefits of VR services may be found by excluding *diagnosis & evaluation* services from the benefits computation. As noted in Section 5, the estimated long-run benefits associated *diagnosis & evaluation* services is likely to be downward biased. While the extent of this bias is unknown, the best reason to think that they may have a true negative effect on labor market outcomes is that they encourage those who are unlikely to succeed in the labor market to sign up for DI/SSI benefits which, from a social welfare point of view, should not be thought of as a negative benefit. Thus, setting the benefits of *diagnosis & evaluation* to zero will lead to a conservative and more accurate estimate of the long-run benefits. In this scenario, we estimate that the mean discounted benefits are $4233 with a standard deviation of $3678 using the 5-year window and $8374 with a standard deviation of $7744 using a 10-year window.

While these estimated benefits are derived directly from the structural model, there are several reasons they may not reflect the true social benefits of VR services. First, some of the estimated earning benefits may reflect the displacement of non-VR participants, particularly if VR services do not improve the VR participant skills or the job matching process. In general, however, training programs for low-skilled workers are not thought to cause notable labor market displacements (see Lalonde, 1995). Second, VR services may lead to other social benefits associated with the increased attachment to the labor market and the resulting reduction in use of the social welfare system. While society does not benefit from reduced transfer payments or increased tax revenues – taxpayer gains exactly offset VR participant losses (except for changes in deadweight loss) – social benefits may result from reduced administrative cost associated with welfare programs and increased VR participant utility due to reduced welfare dependence (Lalonde, 1995), improved health status, and access to health care insurance. At the same time, the deadweight costs of taxations may change if welfare receipt and tax payments change. Likewise, we do not include the estimated positive impact of VR services on DI/SSI receipt. Finally, there is substantial heterogeneity in the discounted benefits across the VR participants,

---

[39]This simulation has a similar structure to the one used to compute marginal effects in Section 5.1 (see Figure 9). But here we compute the present discounted value of the actual treatments provided by DARS rather than a conjectured treatment for single service, j. Formally, we first compute the short- and long-run effect of the program for each individual:

$$\Delta_i = v_{ik}(y_i) - v_{ik}(0)$$

where $v_{ik}(y_i)$ is the estimated labor market earnings under the realized services y_i and $v_{ik}(0)$ is the estimated earnings that would be observed if no services were provided.

[40]Expected earnings are extrapolated in years 9 and 10.

Table 13: Conditional Moments of Expenditure Data on
Purchased VR Services for the 2000 Applicant Cohort

Service	Mean Conditional Expenditure	Std Dev	% with Positive Expenditures
Diagnosis & Evaluation	$747	$756	36%
Training	$2,855	$2,776	41%
Education	$733	$752	2%
Restoration	$361	$770	21%
Maintenance	$441	$1,149	25%
Other Services	$226	$210	3%
Total	$2,418	$2,933	67%

Notes: Moments do not include cost of in-house services or similar benefits.

suggesting that there may be a great deal of variation in the overall benefits estimates (see Figure 9). Many of the clients are estimated to have negative benefits from VR services.

As noted in Section 3, DARS services are provided in any combination of three ways: a) internally by DARS personnel, b) as a "similar benefit" (i.e., purchased or provided by another governmental agency or not-for-profit organization with no charge to DARS), and/or c) as a purchased service through an outside vendor using DARS funds. The DARS data report purchased services but not in-house services or "similar benefits." Table 13 displays the mean costs of purchased services for each service. Interestingly, the average cost for training is substantially less than the mean long-term discounted marginal benefit of $7700 (see Figure 9). Overall, the mean costs of purchased services among all 1260 clients who use purchased services equals $1903 with a standard deviation of $1372. These mean cost estimates have not been discounted, and thus will be inflated to the extent the purchased services are provided over long periods.

The DARS data do not provide information on the costs other than purchased services (missing are DARS-provided services, "similar benefits," and the cost of administrating the program). To estimate these costs, we use information on DARS spending by fiscal year as reported to the US Social Security Administration. These reports summarize information on aggregate administrative costs, DARS-provided counseling, guidance, and placement service costs, purchased service cost, WWRC costs, and size of the caseload for each fiscal year. Except for the WWRC reports, these reports do not provide information on the costs associated with "similar benefits."[41] While there is some variation in the distribution of costs across years, in general, non-purchased service and administrative costs account for 45% of total expenditures, reflecting an average cost per client of roughly $200 per month.

While these reports do not provide information specific to the different impairment groups, this auxiliary information can be used to infer the cost for

---

[41] The fraction of clients estimated to receive (non-WWRC) similar benefits range from between 9% for other services to 16% for education services, but in general these services are provided in combination with DARS purchased services. The fraction of clients only receiving similar benefits varies from 3-4% for *diagnosis & evaluation*, *training*, and *restoration*, to 5-6% for *maintenance* and *other*, and to 8% for *education*.

our sample of applicants with mental illnesses. Two different approaches are used. In the first, we anchor on the fact that purchased services account for 55% of total VR costs. Given that purchased service costs for our sample average $1903 per client receiving purchased services, fixed costs are estimated to be $2325 ($\approx$ ($1903/0.45) $-$ $1903) per client. In the second, we anchor on the fact that the average costs of administration and non-purchased services is $200 per client-month. Given that the average service spell length is 6 quarters, these costs are estimated to be $3600 ($= 3 * 6 * 200$) per client. These two estimates reflect our uncertainty about the costs of non-purchased services and administration. Cases of individuals with mental illness may differ from the general population in both average purchased service expenditures and average spell lengths. So, if cases of individuals with mental illness have low average purchased services relative to non-purchased service costs, the first approach would be downward-biased. If instead, such cases have relatively low average costs associated with administration or non-purchased services, the second approach would be upward-biased. Finally, note that we do not compute separate estimates based on client-specific information on purchased services and spell length. We choose to use only an average "fixed" cost because the model and estimation procedure used to infer benefits allows neither service duration nor actual expenditures to affect labor market outcomes.

Comparing these estimated costs and benefits reveals that DARS services provided to mentally ill people have a substantial positive return especially in the longer run. In total, our preferred estimates, which exclude *diagnosis & evaluation*, imply that mean benefits range from $4233 for the short run to $8374 for the long run, while mean costs range from $4300 to $5500. Thus, even under the most conservative assumptions about the costs of services, the long-run social benefit is estimated to exceed costs by 67%.

We also can compute the rate of return for each person receiving services in our sample. The results of this exercise are reported in Figure 12. For each sample individual receiving some service, we compare the expected flow of benefits they would get with the service package they received relative to the flow of benefits they would get with no services. We approximate cost as[42]

$$f + \sum_{j=1}^{J} y_{ij} c_j$$

where f is a combination of administrative costs and average (unobserved) in-house service and "similar benefits" costs, y_{ij} is an indicator for receipt of service j by person i (as defined in equation (1)), and c_j is the average cost associated with service j computed as the ratio of "mean expenditure" and "% with positive expenditure" in Table 13.

---

[42] An alternative is to use actual cost for each individual. The attractive feature of such an approach is that there is significant variation in cost even conditional on the set of services received. However, we choose to use only average costs for each service because, in the model and estimation procedure, we do not allow actual expenditures to affect labor market outcomes.

Figure 12 shows the distribution of quarterly rates of return for six scenarios: three with $f = \$2400$ and three with $f = \$3600$; and, for each assumption about f, we consider a) a 10-year horizon excluding *diagnosis & evaluation*, b) a 10-year horizon including *diagnosis & evaluation*, and c) a 5-year horizon excluding *diagnosis & evaluation*.[43] Two general lessons emerge from this figure. First, it is clear that earnings flows in years 6 through 10 have a significant impact on estimated rates of return, at least for conventional rates of return.[44] Thus, it is important to use long panels of earnings data such as ours when estimating rates of return.[45] Second, it is clear that exclusion of *diagnosis & evaluation* has a significant impact on the distribution. As noted above, our preferred estimates, which are arguably downward biased, exclude *diagnosis & evaluation*.

The figure also illustrates wide variation in the rates of return across individuals. Focusing on the distribution curve associated with a 10-year horizon and excluding *diagnosis & evaluation*, the distribution curve shows that 19.2% of clients with mental illness have negative rates of return if $f = \$2400$, and 30.5% have negative rates of return if $f = \$3600$ (i.e., there is no positive discount rate that will justify the cost of services relative to the flow of future benefits). At the same time, even if $f = \$3600$, the median rate of return is quite high at 1.9% quarterly (7.9% annually), and 10% of rates of return are above 7.4% quarterly (32.8% annually); if $f = \$2400$, the median rate of return is 3.1% quarterly (12.9% annually), and 10% of rates of return are above 8.8% quarterly (40.2% annually). Meanwhile, including *diagnosis & evaluation* in the analysis causes the proportion with negative returns to increase significantly (for $f = \$3600$, it increases from 30.5% to 58.9%). Likewise, the proportion with negative returns increases significantly when focusing on the distribution curves associated with a 5-year horizon. It should be noted that the variation in rates of return here are due solely to variation in observable characteristics of individuals and variation in the set of services they receive; it is not due to randomness inherent in labor market experience.

Earlier, in Section 5.4, Figure 11 showed that the employment probabilities are overestimated after VR service provisions and underestimated before service. Together, these imply that we might be overestimating the effect of VR services on employment rates. Consider, for example, the distribution of quarterly rates of return for the 10-year horizon without *diagnosis & evaluation* and with estimated fixed costs of $2400. Adjusting all of the predicted employment probabilities by the bias reported in Figure 11 causes the proportion of people

---

[43]Note that, when excluding *diagnosis & evaluation*, we a) ignore observations receiving only *diagnosis & evaluation* and b) ignore all costs and benefits associated with receipt of *diagnosis & evaluation*.

[44]At very high rates of return, later years become irrelevant because of the implied heavy discounting. For example, at a 20% quarterly rate of return, the discount factor associated with earnings 6 years in the future is 0.013.

[45]Estimated rates for returns for non-VR government training programs aimed at economically disadvantaged people also tend to be sensitive to short versus long horizons, and vary widely across programs, demographics, and studies. In some cases, these training programs are found to have average rates of return that are negative. But, in many others, the average annual rates of return are in excess of 100% (Friedlander, 1997; and LaLonde, 1995).
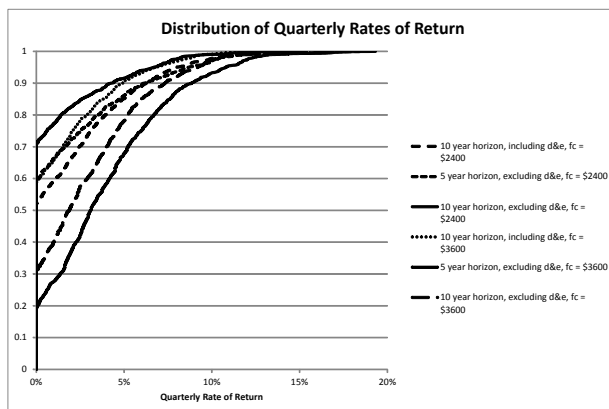
Figure 12: Distribution of Quarterly Rates of Return

with negative rates of return to increase. In particular, the proportion of people with negative rates of return increases from 19.2% to 27.0%, the median quarterly return decreases from 3.1% to 2.0% (8.4% annual rate of return), and 10% of quarterly rates of return still are 7.0% (31.0% annual rate of return). However, this bias correction should be interpreted as suggestive because it does not control for other sources of bias and because we have not computed the standard error of the bias correction.

7 Conclusions

Recently, there have been a number of state-level return on investment evaluations of VR services produced by economic consulting firms or university research bureaus (e.g., Heminway and Rohani, 1999; Uvin, Karaaslani, and White, 2004; Hollenbeck and Huang, 2006; Kisker et al., 2008; and Wilhelm and Robinson, 2010).[46] By comparing outcomes of a "treated" and "untreated" group, as we do in Figures 4 and 5, these studies tend to find large positive returns to VR services. An evaluation of Utah's VR program, for example, found that the public benefits of the program, measured in dollars, exceed the cost by a factor of 5.64 (Wilhelm and Robinson, 2010). These reports, however, have a number of serious shortcomings which are addressed in this paper, including a) identification problems; b) problems caused by censored data; c) the selection problem; and d) heterogeneity in the caseload and in the services provided. Our analysis of the Virginia VR program addresses important limitations of these

---

[46]These state level studies condition the analysis on observed covariates. In some cases (Hollenbeck and Huang, 2006), researchers use statistical matching estimators based on propensity scores, initially developed by Rosenbaum and Rubin (1983) and incorporated in other manpower training program evaluations (e.g., Heckman, Ichimura, and Todd,1997; Dehejia and Wahba, 1999). All of these analyses, however, invoke a conditional independence assumption that the outcome is independent of provision of services.

recent studies. First, using the model described in Section 2, we formally account for the possibility that selection into the treatment is endogenous. As noted above, a simple comparison of mean outcomes among treated and untreated clients may be spurious due to selection, and conditioning on observed covariates is not likely to address this problem credibly. Our results suggest that selection plays an important role in inferences on the effect of VR services. Second, by focusing on clients with mental illnesses, we allow the estimated effects of treatment to vary with the clients' limiting conditions. In contrast, these state-level reports do not distinguish between clients with mental illness, cognitive impairments, sensory impairments, or physical impairments. Arguably, the effects of the program are heterogeneous, and restricting the impact to be constant across all groups may lead to biased inferences. Third, unlike these earlier evaluations, we examine the impact of specific types of services rather than just a single treatment indicator. We find that services do, in fact, have very different impacts on labor market effects. Finally, we observe labor market and disability insurance outcomes many years before and after the provision of VR services. In this analysis, being able to estimate the long-run return is critical as it significantly differs from the short-run return.

Our results suggest a complex picture of the impact of VR services on labor market outcomes and DI/SSI receipt. Pre-program labor market differences vary across the six service types, estimated employment effects are positive for some services (e.g., *training*) and negative for others (e.g., *education*), and estimated earnings effects are consistently positive. When combining the employment and earnings effects together, we find that, except for *diagnosis & evaluation*, all of the other service types have positive long-run effects. On average, *training*, *restoration*, and *other services* have benefits on the order of $7200, $3750, and $4800 respectively, while *education* and *maintenance* have positive benefits of $1700 and $2100 respectively. Overall, we find that VR services have a positive average return, with mean long-run benefits of $4124 or $8374, depending upon how one interprets *diagnosis & evaluation* results, and mean costs between $3460 and $5060. We also find, however, much variation in the return across VR participants. Depending upon how one estimates fixed costs f (and excluding *diagnosis & evaluation*), between 11.5% ($f = \$1560$) and 30.5% ($f = \$3600$) of VR participants with mental illness have negative long-run rates of return, half have long-run annual rates of return in excess of between 17.5% ($f = \$1560$) and 7.7% ($f = \$3600$), and 10% have annual long-run annual rates of return in excess of between 50.7% ($f = \$1560$) and 32.8% ($f = \$3600$). Finally, our results suggests that VR programs are unlikely to reduce the burgeoning growth in DI/SSI roles. To the contrary, we find that these services increase the probability VR clients take-up DI/SSI.

40

8 Appendices

8.1 Covariance Structure

The covariance matrix of the errors $u_i' = (u_{i1}^y, u_{i2}^y, .., u_{iJ}^y, u_{i1}^z, u_{i1}^w, u_{i1}^r .., u_{iT}^z, u_{iT}^w, u_{iT}^r)$ implied by the structure in equation (5) of the paper is

$$
\Omega_{(J+3T)\times(J+3T)} = \left(\begin{array}{cc} \underset{J\times J}{A} & \underset{J\times 3T}{B'} \\ \underset{3T\times J}{B} & \underset{3T\times 3T}{C} + \underset{3T\times 3T}{D} \end{array} \right)
$$

where

$$
A = \left(\begin{array}{cccc} \sum_k \left(\lambda_{1k}^y\right)^2 & \sum_k \lambda_{1k}^y\lambda_{2k}^y & \cdots & \sum_k \lambda_{1k}^y\lambda_{Jk}^y \\ \sum_k \lambda_{1k}^y\lambda_{2k}^y & \sum_k \left(\lambda_{2k}^y\right)^2 & \cdots & \sum_k \lambda_{2k}^y\lambda_{Jk}^y \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \lambda_{1k}^y\lambda_{Jk}^y & \sum_k \lambda_{2k}^y\lambda_{Jk}^y & \cdots & \sum_k \left(\lambda_{Jk}^y\right)^2 \end{array} \right),
$$

$$
C = H \otimes Q_T
$$

$$
\underset{3\times 3}{H} = \left(\begin{array}{ccc} \sum_k \left(\lambda_k^z\right)^2 & \sum_k \lambda_k^z\lambda_k^w & \sum_k \lambda_k^z\lambda_k^r \\ \sum_k \lambda_k^z\lambda_k^w & \sum_k \left(\lambda_k^w\right)^2 & \sum_k \lambda_k^w\lambda_k^r \\ \sum_k \lambda_k^z\lambda_k^r & \sum_k \lambda_k^w\lambda_k^r & \sum_k \left(\lambda_k^r\right)^2 \end{array} \right),
$$

$$
\underset{T\times T}{Q_T} = \left(\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{array} \right)
$$

$$
D = \underset{3\times 3}{\Omega_\zeta} \otimes \frac{1}{1-\rho_\eta^2} \left(\begin{array}{cccc} 1 & \rho_\eta & \cdots & \rho_\eta^{T-1} \\ \rho_\eta & 1 & \cdots & \rho_\eta^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_\eta^{T-1} & \rho_\eta^{T-2} & \cdots & 1 \end{array} \right),
$$

and

$$
B = q_T \otimes F,
$$

$$
\underset{T\times 1}{q_T} = \left(\begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array} \right),
$$

$$
\underset{3\times J}{F} = \left(\begin{array}{cccc} \sum_k \lambda_{1k}^y\lambda_{1k}^z & \sum_k \lambda_{2k}^y\lambda_{1k}^z & \cdots & \sum_k \lambda_{Jk}^y\lambda_{1k}^z \\ \sum_k \lambda_{1k}^y\lambda_{1k}^w & \sum_k \lambda_{2k}^y\lambda_{1k}^w & \cdots & \sum_k \lambda_{Jk}^y\lambda_{1k}^w \\ \sum_k \lambda_{1k}^y\lambda_{1k}^r & \sum_k \lambda_{2k}^y\lambda_{1k}^r & \cdots & \sum_k \lambda_{Jk}^y\lambda_{1k}^r \end{array} \right).
$$

8.2 Counselor and Field Office Effects

We use as an instrument in equation (1) of the paper, a transformation of the proportion of other clients of the same counselor provided service j, i.e., a counselor effect. We also use a transformation of the proportion of other clients from the same office provided service j, i.e., an office effect. We transform the counselor and office effects using an inverse normal distribution function to make it more likely that, as the counselor and office effects vary, their effect on service probabilities can vary by approximately the same amount. To consider why this is attractive, consider a counselor who almost always uses a particular service. We want to allow for the possibility that this will imply that all of the clients of the counselor are very likely to receive that service. Limiting the counselor effects to vary between $(0, 1)$ makes it harder for that to occur. On the other hand, using an inverse distribution function for a distribution with the real line as support makes the range $(-\infty, \infty)$.

While such a transformation makes sense analytically, in practice, it might cause problems for values of the untransformed effect at or near the boundaries. We propose a "fix" that both makes sense and solves the boundary problem. In particular, we propose replacing the untransformed effect c_{ij} with

$$c_{ij}^* = (1 - \omega_i) c_{ij} + \omega_i \bar{c}_j \tag{12}$$

where $\bar{c}_j$ is the mean value of c_{ij} across all counselors (offices), $\omega_i = \kappa_i^{-1}$, and κ_i is the number of clients seen by counselor i (office i). This specification allows the counselor effect and office effect to be more important for those counselors (offices) who have many observed clients. In fact, it has a certain Bayesian flavor to it.

There are some respondents who either have missing counselor or office information or who have a counselor (or office) with no other clients. For such cases, we can not create our effects.[47] Because of such cases, we include a set of dummies for missing counselor and/or missing office effects. It turns out that these dummies are very highly correlated, and most of the missing office effects must be excluded from the model to avoid a singular Hessian.

Tables A.1 and A.2 provide information about the moments of the transformed counselor and office effects. One can see that there is significant variation in both. There is some evidence of left-tailed skewness but no unreasonable outliers. The lack of outliers occurs despite zeroes for some services for some counselors and field offices because of the weighted average inherent in equation (12).

8.3 Local Labor Market Conditions

Virginia is unique among states in that it has both counties and independent cities. While BEA provides data for almost all counties and independent cities, there is a small number of mostly rural counties for which BEA provides data

---

[47] In fact, when a counselor (office) has only one other client, we treat it as missing also.

Table A.1: Moments of Inverse Normal Transformed Office Effects

Service	Mean	Std Dev	Minimum	Maximum
Diagnosis & Evaluation	-0.396	0.264	-1.512	0.785
Training	-0.149	0.255	-1.133	0.707
Education	-1.200	0.388	-2.699	-0.189
Restoration	-0.719	0.463	-2.469	0.511
Maintenance	-0.502	0.410	-1.754	0.419
Other Service	-0.828	0.595	-2.668	0.405

Note: # Obs = 1489.

Table A.2: Moments of Normal Logistic Transformed Counselor Effects

Service	Mean	Std Dev	Minimum	Maximum
Diagnosis & Evaluation	-0.412	0.424	-2.061	1.045
Training	-0.173	0.513	-1.795	1.472
Education	-1.351	0.625	-2.542	0.66
Restoration	-0.805	0.615	-2.298	0.735
Maintenance	-0.549	0.564	-2.105	0.802
Other Service	-0.883	0.697	-2.303	1.054

Note: # Obs = 1485.

Table A.3: Aggregated Regions

Region	Component Counties	Component Independent Cities
Eastern Shore	Accomack, Northhampton	
Rural Shenandoah, South	Bath, Highland, Rockbridge	Alleghany, Covington, Buena Vista, Lexington
Dinwiddie	Brunswick, Lunenberg, Nottoway	Dinwiddie, Colonial Heights, Petersburg, Greensville, Emporia
Bluefield, WV-VA Micropolitan SA	Bland, Buchanan, Dickerson, Lee, Norton, Smyth, Tazewell, Wythe, Wise	
Lynchburg Rural	Buckingham, Prince Edward	
Danville Rural	Charlotte, Halifax, Mecklenburg	
Northern Neck	Essex, King George, Lancaster, Middlesex, Northumberland, Richmond, Westmoreland	
Martinsville Rural	Floyd, Grayson, Patrick	Carroll, Galax
Culpeper, VA Micropolitan SA	Culpeper, Madison, Orange, Rappahannock	
Franklin/ Southhampton	Franklin, Southhampton	
Harrisonburg Rural	Page, Shenandoah	

only after some aggregation. We create 11 aggregated regions to deal with this problem listed in Table A.3.

We construct the employment rate by dividing number of people employed by working age population. We do this both at the county/independent city level and at the MSA level.[48] Significant variation in these measures exists across time, across geography, and across the two separate measures. One should note that there are some counties with employment rates greater than one. This occurs because the population numbers are based on county of residence while the employment numbers are based on county where one works. Thus, these rates reflect variation in net commuting patterns across counties.

---

[48] In the paper, we use only the county/independent city level because the two measures are very highly correlated.

8.4 Bias Caused by Unobserved Heterogeneity in Measured Mental Illness

There are many possibilities to explain the results with respect to the *diagnosis & evaluation* effects, including, but not limited to, the possibility that a) the instrument is correlated with the errors; b) the estimated net effect of *diagnosis & evaluation* on long-run outcomes is negative is a statistical anomaly and one might reject the (one-sided) null hypothesis that all of the long-run outcomes are positive; and c) extra diagnosis and evaluation requires much time for people with mental illness and thus slows down the rehabilitation process (note that the negative effect is due solely to employment effects).The problem with (a) is that the result is specific to *diagnosis & evaluation*, and it disappears for other disability groups (e.g., see Dean, et al., 2013). The problem with (b) is that the null hypothesis would be rejected. We have no information on (c).

The bias explanation we prefer, which is also confirmed by DARS counselors, is the explanation included in the text of the paper. The idea is that, for people with mental illness, receipt of purchased diagnosis & evaluation services is an indicator that the individual's mental health problem is particularly difficult to deal with in a way unobserved in the DARS administrative data. This unobserved heterogeneity in mental illness is an error in measurement of an explanatory variable, and it causes *diagnosis & evaluation* to be correlated with the errors in the labor market outcomes equations. More explicitly, but in a simpler linear context, consider the model,

$$y_i = X_i\beta + w_{1i}\alpha_1 + w_{2i}\alpha_2 + u_i$$

where y_i is an outcome variable of interest for observation i, X_i is a vector of exogenous explanatory variables, w_{1i} is a potentially endogenous explanatory variable such as receipt of *diagnosis & evaluation*, and w_{2i} is another explanatory variable measured with error such as degree of mental illness with

$$plim\left(n^{-1}\sum_i w_{1i}w_{2i}\right) > 0.$$

In particular, for simplicity, assume that

$$w_{2i} \in \{0, 1, 2\},$$

but w_{2i} is not observed, and, instead,

$$x_i = 1\,(w_{2i} > 0)$$

is observed. Then the model to be estimated

$$y_i = X_i b + w_{1i}a_1 + x_i a_2 + v_i.$$

Let Z be a matrix of instruments. Then the asymptotic properties of the IV

estimator are

$$plim \begin{pmatrix} \widehat{b} \\ \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = plim \begin{pmatrix} Z'X/n \\ Z'w_1/n \\ Z'x/n \end{pmatrix}^{-1} plim\left(Z'y/n\right)$$

$$= plim \begin{pmatrix} Z'X/n \\ Z'w_1/n \\ Z'x/n \end{pmatrix}^{-1} \left[plim \begin{pmatrix} Z'X/n \\ Z'w_1/n \\ Z'w_2/n \end{pmatrix} \begin{pmatrix} \beta \\ \alpha_1 \\ \alpha_2 \end{pmatrix} + Z'u/n \right]$$

$$= plim \begin{pmatrix} Z'X/n \\ Z'w_1/n \\ Z'x/n \end{pmatrix}^{-1} plim \begin{pmatrix} Z'X/n \\ Z'w_1/n \\ Z'w_2/n \end{pmatrix} \begin{pmatrix} \beta \\ \alpha_1 \\ \alpha_2 \end{pmatrix} \neq \begin{pmatrix} \beta \\ \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Now, in the interest of making more progress, consider a special case where

$$w_{1i} = \gamma_0 + \gamma_1 w_{2i} + e_i.$$

Then

$$plim \begin{pmatrix} \widehat{b} \\ \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = = plim \begin{pmatrix} Z'X/n \\ Z'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ Z'\left[1\left(w_2 > 0\right)\right]/n \end{pmatrix}^{-1}.$$

$$plim \begin{pmatrix} Z'X/n \\ Z'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ Z'w_2/n \end{pmatrix} \begin{pmatrix} \beta \\ \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Next, in the same spirit, assume that $\beta = 0$; i.e., there are no X's (wo/ this assumption, as is the case in any measurement error problem, the sample correlation of X with w_1 and w_2 contaminates the analysis relative to the simpler case). Then

$$plim \begin{pmatrix} \widehat{a}_1 \\ \widehat{a}_2 \end{pmatrix} = plim \begin{pmatrix} z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n & z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ z_1'x/n & z_2'x/n \end{pmatrix}^{-1}.$$

$$plim \begin{pmatrix} z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n & z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ z_1'w_2/n & z_2'w_2/n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$= \frac{plim \begin{pmatrix} z_2'x/n & -z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ -z_1'x/n & z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \end{pmatrix}}{plim\left[\left(\frac{z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right)\left(\frac{z_2'x}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right)\right]}.$$

$$plim \begin{pmatrix} z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n & z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]/n \\ z_1'w_2/n & z_2'w_2/n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$= plim D^{-1} plim \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

where

$$A_{11} = \left(\frac{z_2'x}{n}\right)\left(\frac{z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right) - \left(\frac{z_1'w_2}{n}\right)\left(\frac{z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right),$$

$$A_{12} = \left[\left(\frac{z_2'x}{n}\right) - \left(\frac{z_2'w_2}{n}\right)\right]\left(\frac{z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right),$$

$$A_{21} = \left[\left(\frac{z_1'w_2}{n}\right) - \left(\frac{z_1'x}{n}\right)\right]\left(\frac{z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right),$$

$$A_{22} = \left(\frac{z_2'w_2}{n}\right)\left(\frac{z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right),$$

$$D = \left[\left(\frac{z_2'x}{n}\right)\left(\frac{z_1'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'\left[\gamma_0 1 + \gamma_1 w_2 + e\right]}{n}\right)\right].$$

Note that, in the case where $x = w_2$ (i.e., there is no measurement error),

$$plim\left(\begin{array}{c}\widehat{a}_1\\\widehat{a}_2\end{array}\right) = \left(\begin{array}{cc}1 & 0\\0 & 1\end{array}\right)\left(\begin{array}{c}\alpha_1\\\alpha_2\end{array}\right) = \left(\begin{array}{c}\alpha_1\\\alpha_2\end{array}\right).$$

In general,

$$plim\left(\begin{array}{c}\widehat{a}_1\\\widehat{a}_2\end{array}\right) = \left(\begin{array}{cc}1 & \eta_{12}\\\eta_{21} & 1\end{array}\right)\left(\begin{array}{c}\alpha_1\\\alpha_2\end{array}\right)$$

where

$$\eta_{12} = \frac{plim\left[\left(\frac{z_2'x}{n}\right) - \left(\frac{z_2'w_2}{n}\right)\right]\left(\frac{z_2'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right)}{plim\left[\left(\frac{z_2'x}{n}\right)\left(\frac{z_1'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right)\right]},$$

$$\eta_{21} = \frac{plim\left[\left(\frac{z_1'w_2}{n}\right) - \left(\frac{z_1'x}{n}\right)\right]\left(\frac{z_1'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right)}{plim\left[\left(\frac{z_2'x}{n}\right)\left(\frac{z_1'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'[\gamma_0 1 + \gamma_1 w_2 + e]}{n}\right)\right]}.$$

Without loss of generality, we can assume that

$$plim\left(\frac{z_1'1}{n}\right) = plim\left(\frac{z_2'1}{n}\right) = plim\left(\frac{z_1'e}{n}\right) = plim\left(\frac{z_2'e}{n}\right) = 0;$$

$$plim\left(\frac{z_2'w_2}{n}\right) > 0$$

which implies that

$$\eta_{12} = \frac{plim\left(\frac{z_2'(x - w_2)}{n}\right)\left(\frac{z_2'w_2}{n}\right)}{plim\left[\left(\frac{z_2'x}{n}\right)\left(\frac{z_1'w_2}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'w_2}{n}\right)\right]};$$

$$\eta_{21} = \frac{plim\left(\frac{z_1'(w_2 - x)}{n}\right)\left(\frac{z_1'w_2}{n}\right)}{plim\left[\left(\frac{z_2'x}{n}\right)\left(\frac{z_1'w_2}{n}\right) - \left(\frac{z_1'x}{n}\right)\left(\frac{z_2'w_2}{n}\right)\right]}.$$

47

Effects of Client Characteristics on Service Receipt by Type

Variable	Diagnosis & Evaluation	Training	Education
Constant	0.631	1.199 *	-1.183
Male	-0.102	-0.315 *	-0.118
White	-0.056	-0.077	-0.251
Education	-0.011	-0.005	0.082 **
Special Education	0.458	-0.529	0.125
Education Missing	-1.008 **	-4.000 +	-2.583 *
Age/100	0.143	-0.247	-0.110
Married	-0.145	-0.299	-0.019
# Dependents	-0.044	-0.198 **	-0.095
Transportation Available	0.182	0.152	0.493 *
Has Driving License	0.231	-0.252	0.679 **
Receives Govt Assistance	0.491 **	0.777 **	0.313
Musculoskeletal Disability	0.241	-0.489 **	0.009
Learning Disability	0.680 **	0.074	-0.628
Mental Illness	-0.640 **	-0.976 **	-0.464
Substance Abuse	0.027	-0.373 *	0.265
Disability Significant	0.350 *	0.136	-0.355
Disability Most Significant	0.466 *	0.502	-0.406
SMI	0.028	0.714	0.350
Male * SMI	-0.146	0.041	-0.664
White * SMI	-0.081	0.153	0.448
Education * SMI	0.005	-0.063	0.008
Age/100 * SMI	-0.062	0.316	-0.443

Effects of Client Characteristics on Service Receipt by Type (continued)

Variable	Restoration	Maintenance	Other Services
Constant	0.860 *	1.672 **	0.113
Male	-0.301 *	0.038	-0.095
White	-0.155	-0.393 *	-0.283
Education	-0.008	-0.032	0.011
Special Education	-0.020	-0.921 *	-0.124
Education Missing	-1.386 **	-4.000 +	-2.429 **
Age/100	-0.039	-0.386 *	0.010
Married	-0.254	-0.405 *	-0.235
# Dependents	0.095	-0.015	-0.134 *
Transportation Available	-0.066	-0.149	0.382 *
Has Driving License	0.142	-0.337 *	0.058
Receives Govt Assistance	-0.119	0.378	0.101
Musculo/Skeletal Disability	-0.134	0.128	-0.004
Learning Disability	2.071 **	0.036	0.556 *
Mental Illness	-0.432	-0.534	-0.592 *
Substance Abuse	-0.103	0.329 *	-0.192
Disability Significant	0.095	0.022	0.004
Disability Most Significant	0.335	0.396	0.172
SMI	-1.441	-0.558	0.469
Male * SMI	0.228	-0.277	0.042
White * SMI	0.030	0.231	0.088
Education * SMI	0.028	-0.047	-0.031
Age/100 * SMI	-0.123	0.818 *	-0.053

Notes:

1. Standard errors not presented to save space but are available from the corresponding author.

2. Single-starred items are statistically significant at the 10% level, and double-starred items are statistically significant at the 5% level, and items with + were restricted.

If the denominator is positive (z_2 close to w_2 and z_1 close to w_1) and $plim\left(\frac{z_2'(x-w_2)}{n}\right)$ is a better instrument for w_2 than for x (note that these assumptions would all be true if ($z_2 = w_2$ or $z_2 = x$) and $z_1 = w_1$), then
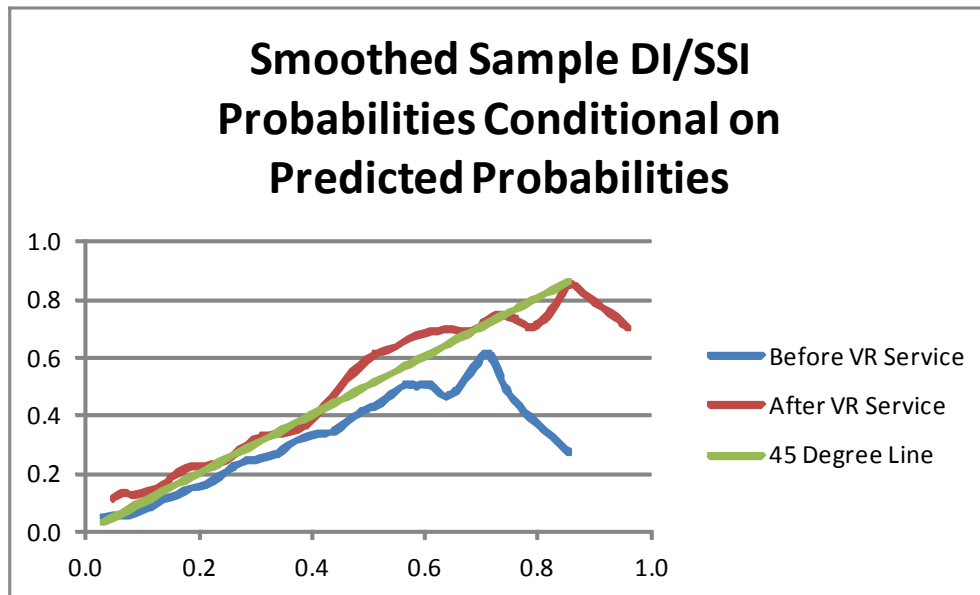
$$\eta_{12} < 0, \eta_{21} > 0,$$

which implies that

$$plim\,\widehat{a}_1 \quad < \quad \alpha_1,$$
$$plim\,\widehat{a}_2 \quad > \quad \alpha_2.$$

In words, the estimate on diagnosis & evaluation would be negatively biased, and the estimate on *mental illness* or *SMI* would be biased upwards.

8.5 Effects of Client Characteristics on Service Receipt by Type

8.6 Smoothed Sample DI/SSI Probabilities

Smoothed Sample DI/SSI Probabilities Conditional on Predicted Probabilities

Legend: Before VR Service, After VR Service, 45 Degree Line

References

[1] Aakvik, A., J. Heckman, and E. Vytlacil (2005). "Estimating Treatment Effects For Discrete Outcomes When Responses To Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs." *Journal of Econometrics.* 125: 15-51.

[2] Arrighi, Y., B. Davin, A. Trannoy, and B. Ventelou (2010). "Un Effet Prix dans la Demande d'APA." Unpublished manuscript.

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

[4] Autor, David and Mark Duggan (2010). *Supporting Work: A Proposal for Modernizing the U.S. Disability Insurance System.* Hamilton Project.

[5] Baldwin, Marjorie (1999). "The Effects of Impairments on Employment and Wages: Estimates from the 1984 and 1990 SIPP." *Behavioral Sciences and the Law.* 17: 7-27.

[6] Baldwin, Marjorie (2005). "Explaining the Differences in Employment Outcomes Between Persons With and Without Mental Disorders." Unpublished manuscript.

[7] Bassi, L. (1984). "Estimating the Effect of Training with Non Random Selection." *Review of Economics and Statistics.* 66: 36-43.

[8] Bell, S., L. Orr, J. Blomquist, and G. Cain (1995). *Program Applicants as a Comparison Group in Evaluating Training Programs*. Upjohn Institute for Employment Research.

[9] Bellante, D. (1972). "A Multivariate Analysis of a Vocational Rehabilitation Program." *Journal of Human Resources*. 7: 226-241.

[10] Benitez-Silva, Hugo, Moshe Buchinsky, Hiu Man Chan, John Rust, and Sofia Sheidvasser. (1999). "An Empirical Analysis of the Social Security Application, Appeal and Award Process." *Labour Economics*. 6: 147-178.

[11] Berkowitz, E. (1988). "The Cost-Benefit Tradition in Vocational Rehabilitation." *Measuring the Efficiency of Public Programs: Costs and Benefits in Vocational Rehabilitation*. (ed.) E. Berkowitz. Temple University Press, 10-28.

[12] Bewley, Helen, Richard Dorsett, and Sergio Salis (2007). *The Impact of Pathways to Work on Work, Earnings and Self-Reported Health in the April 2006 Expansion Areas*. Research Report No 601. Department for Work and Pensions: London.

[13] Bond, G., R. Drake, and D. Becker (2001). "An Update on Randomized Controlled Trials of Evidence-Based Supported Employment." *Psychiatric Rehabilitation Journal*. 31(4): 280-290.

[14] Börsch-Supan, Axel and Vassilis Hajivassiliou (1992). "Health, Children, and Elderly Living Arrangements: A Multiperiod-Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors." *Topics in the Economics of Aging*. (ed.) David Wise. Chicago and London: U. of Chicago Press, 79-104.

[15] Bratberg, E., A. Grasdal, and A. Risa. (2002). "Evaluating Social Policy by Experimental and Nonexperimental Methods." *Scandinavian Journal of Economics*. 104(1): 147-171.

[16] Bureau of Economic Analysis (2010a). http://www.bea.gov/regional/docs/footnotes.cfm?tablename=CA04.

[17] Bureau of Economic Analysis (2010b). http://www.bea.gov/regional/docs/footnotes.cfm?tablename=CA25.

[18] Burns, Tom, Jocelyn Catty, Thomas Becker, Robert Drake, Angelo Fioritti, Martin Knapp, Christoph Lauber, Wulf Rössler, Toma Tomov, Jooske van Busschbach, Sarah White, Durk Wiersma, Connie Geyer, Marsha Koletsi, Pascale Lissouba, Miles Rinaldi, Ulrike Ehiosun, Rana Kalkan, Reinhold Kilian, Denise Manchisi, Astrid Niersman, Ingeborg Warnke, Dimitar Germanov, Adelina Comas, Claire Curran, and Anita Patel (2007). "The Effectiveness of Supported Employment for People with Severe Mental Illness: A Randomized Controlled Trial." *Lancet*. 370(9593): 1146-1152.

[19] Butler, Richard, William Johnson, and Marjorie Baldwin (1995). "Managing Work Disability: Why First Return to Works is Not a Measure of Success." *Industrial and Labor Relations Review.* 48(3): 452-469.

[20] Campbell, Kikuko, Gary Bond, Robert Drake, Gregory McHugo, and Haiyi Xie (2010). "Client Predictors of Employment Outcomes in High-Fidelity Supported Employment: A Regression Analysis." *Journal of Nervous and Mental Disease.* 198(8): 556-563.

[21] Cimera, R. (2010). "The National Cost-Efficiency of Supported Employees with Intellectual Disabilities: 2002 to 2007." *American Journal on Intellectual and Developmental Disabilities.* 115: 19-29.

[22] Clapp, Christopher, Amalia Miller, John Pepper, and Steven Stern (2010). "Evaluation of Programs for High Risk Youth in Virginia." Unpublished manuscript.

[23] Conley, R. (1969). "A Benefit Cost Analysis of the Vocational Rehabilitation Program." *Journal of Human Resources.* 4: 226-252.

[24] Cook, Judith, Anthony Lehman, Robert Drake, William McFarlane, Paul Gold, H. Stephen Leff, Crystal Blyler, Marcia Toprac, Lisa Razzano, Jane Burke-Miller, Laura Blankertz, Michael Shafer, Susan Pickett-Schenk, and Dennis Grey (2005). "Integration of Psychiatric and Vocational Services: A Multisite Randomized, Controlled Trial of Supported Employment." *American Journal of Psychiatry.* 162(10): 1948-1956.

[25] Currie, Janet and Bridget Madrian (1999). "Health, Health Insurance, and the Labor Market." *Handbook of Labor Economics, Volume 3.* (eds.) Orley Ashenfelter and David Card. Amsterdam : Elsevier.

[26] Dean, David and Robert Dolan (1991). "Fixed-Effects Estimates of Earnings Impacts for the Vocational Rehabilitation Program" *Journal of Human Resources.* 26(2): 380-391.

[27] Dean, David, Robert Dolan, and Robert Schmidt (1999). "Evaluating the Vocational Rehabilitation Program Using Longitudinal Data: Evidence for a Quasi-experimental Research Design." *Evaluation Review.* 23(2):162-189.

[28] Dean, David, Robert Dolan, Robert Schmidt, Paul Wehman, John Kregel, and Grant Revell (2002). "A Paradigm for Evaluation of the Federal-State Vocational Rehabilitation Program." *Achievements and Challenges in Employment Services for People with Disabilities: The Longitudinal Impact of Workplace Supports.* (eds.) John Kregel, David Dean, and Paul Wehman. Richmond: Virginia Commonwealth University Rehabilitation Research and Training Center on Workplace Supports.

[29] Dean, David, John Pepper, Robert Schmidt, and Steven Stern (2013a). "The Effects of Vocational Rehabilitation for People with Cognitive Impairments." Unpublished manuscript.

[30] Dean, David, John Pepper, Robert Schmidt, and Steven Stern (2013b). "The Effects of Youth Transition Programs on Labor Market Outcomes." Unpublished manuscript.

[31] Dehejia, R. and S. Wahba (1999). "Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association.* 94(448): 1053-1062.

[32] Doyle, Joseph (2007). "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." *American Economic Review.* 97(5): 1583-1610.

[33] Drake, Robert, Jonathan Skinner, Gary Bond, and Howard Goldman (2009). "Social Security and Mental Illness: Reducing Disability With Supported Employment." *Health Affairs.* 28(3): 761-770.

[34] Ettner, Susan, Richard Frank, and Ronald Kessler (1997). "The Impact of Psychiatric Disorders on Labor Market Outcomes." *Industrial and Labor Relations Review.* 51(1): 64-81.

[35] French, Eric and Jae Song. (2012). "The Effect of Disability Insurance Receipt on Labor Supply: A Dynamic Analysis." Unpublished Manuscript.

[36] Friedlander, Daniel, David Greenberg, and Philip Robins (1997). "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature.* 35(4), 1809-1855.

[37] Frolich, M., A. Heshmati, and M. Lechner (2004). "A Microeconometric Evaluation of Rehabilitation of Long-Term Sickness in Sweden." *Journal of Econometrics.* 19: 375–396.

[38] Geweke, John (1988). "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference." *Journal of Econometrics.* 38(1/2): 73-89.

[39] Governmental Accountability Office (2005) "Vocational Rehabilitation: Better Measures and Monitoring Could Improve the Performance of the VR Program." Washington, D.C.

[40] Hanushek, Eric, John Kain, and Steven Rivkin (2002). "Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities?" *Review of Economics and Statistics.* 84(4): 584-599.

[41] Heckman, J. and V. 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: 862-874.

[42] Heckman, J., H. Ichimura, and P. Todd (1997). "Matching as an Econometric Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies.* 65: 261-294.

[43] Heckman, James, Robert LaLonde, and Jeffrey Smith (1999). "The Economics and Econometrics of Active Labor Market Programs." *Handbook of Labor Economics Volume 3A*. (eds.) Orley Ashenfelter and David Card. Amsterdam: North-Holland. 1865-2097.

[44] Heckman, James and Burton Singer (1984a). "Econometric Duration Analysis." *Journal of Econometrics*. 24: 63-132.

[45] Heminway, Derek and Faranak Rohani (1999). *A Cost-Benefit Analysis of the Employment of People with Disabilities in Florida: Final Report*. Educational Services Program.

[46] Hennessey, John and L. Scott Muller (1995). "The Effect of Vocational Rehabilitation and Work Incentives on Helping the Disabled-Worker Beneficiary Back to Work." *Social Security Bulletin*. 58(1): 15-28.

[47] Hollenbeck, Kevin and Wei-Jang Huang (2006). *Net Impact and Benefit-Cost Estimates of the Workforce Development System in Washington State*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

[48] Hotz, V. (1992). "Design Analysis Plan for the GAO Study of the Vocational Rehabilitation Program." Correspondence with the GAO Program Evaluation and Methodology Division, 1-10.

[49] Imbens, Guido and Jeffrey Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*. 47(1): 5–86.

[50] Keith, Samuel, Darrel Regier, and Donald Rae (1991). *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. (eds.) Lee Robins and Darrel Regier. MacMillan: New York.

[51] Kessler, R., P. Berglund, M. Bruce, J. Koch, E. Laska, P. Leaf, R. Manderscheid, R. Rosenheck, E. Walters, and P. Wang (2001). "The Prevalence and Correlates of Untreated Serious Mental Illness." *Health Services Research*. 36(6 Pt 1): 987-1007.

[52] Kimmel, Jean and Thomas Kniesner (1998). "New Evidence on Labor Supply: Employment versus Hours Elasticities by Sex and Marital Status." *Journal of Monetary Economics*. 42: 289-301.

[53] Kisker, E., G. Strech, J. Vetter, and C. Foote (2008). *Evaluating Rehabilitation Services in Oklahoma: An Analysis of Program Impacts and of Benefit/Costs*. Oklahoma Department of Rehabilitative Services.

[54] Kreider, Brent (1998). "Workers' Applications to Social Insurance Programs when Earnings and Eligibility are Uncertain." *Journal of Labor Economics*. 16(4): 848-877.

[55] Kreider, Brent (1999). "Social Security Disability Insurance: Applications, Awards, and Lifetime Income Flows." *Journal of Labor Economics.* 17(4): 784-827.

[56] Kreider, Brent and John Pepper (2007). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors." *Journal of the American Statistical Association.* 102(478): 432-41.

[57] LaLonde, Robert (1995). "The Promise of Public Sector-Sponsored Training Programs." *Journal of Economic Perspectives.* 9(2), 149-168.

[58] Loprest, P. (2007). *Strategic Assessment of the State of the Science in Research on Employment for Individuals with Disabilities.* National Institute for Disability and Rehabilitation Research. Contract GS23F8198H.

[59] Maestas, Nicole, Kathleen Mullen, and Alexander Strand (2012). "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." Unpublished manuscript.

[60] Marcotte D., V. Wilxox-Gok, and D. Redmond (2000). "The Labor Market Effects of Mental Illness: The Case of Affective Disorders." *The Economics of Disability.* (eds.) D. Salkever and A. Sorkin. Greenwich, CT: JAI Press.

[61] McFadden, Daniel and Kenneth Train (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics.* 15(5): 447-470.

[62] McKeithen, Amelia and Steven Stern (2007). "The Costs and Benefits of Mental Healthcare." http://www.people.virginia.edu/~sns5r/resint/psychtstf/mentalhealthbenefits.pdf.

[63] Meyer, Bruce (1995). "Natural and Quasi-Experiments in Economics." *Journal of Business and Economic Statistics.* 13(2):151–162.

[64] Mueser, K., M. Salyers, and P. Mueser (2001). "A Prospective Analysis of Work in Schizophrenia." *Schizophrenia Bulletin.* 27: 281-296.

[65] Mueser, Peter, Kenneth Troske, and Alexey Gorislavky (2007). "Using State Administrative Data to Measure Program Performance." *Review of Economics and Statistics.* 89(4): 761-783.

[66] Nowak, L. (1983). "A Cost Effectiveness Evaluation of the Federal/State Vocational Rehabilitation Program Using a Comparison Group." *The American Economist.* 27: 23-29.

[67] Raphael, Steven and Lorien Rice (2002). "Car Ownership, Employment, and Earnings." *Journal of Urban Economics.* 52(1): 109-130.

[68] Raaum, O. and H. Torp (2001). "Labor Market Training in Norway – Effects on Earnings." *Labour Economics.* 8(5): 207-247.

[69] Rogers, J., M. Bishop, and M. Crystal (2005). "Predicting Rehabilitation Outcome for Supplemental Security Income and Social Security Disability Insurance Recipients: Implications for Consideration with the Ticket to Work Program." *Journal of Rehabilitation*. 71(3): 5-10.

[70] Rosenbaum, P. and D. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Casual Effects." *Biometrika*. 70: 41-55.

[71] Sosulski, Marya, Chandra Donnell, and Woo Jong Kim (2012). "Disability and Employee Benefits Receipt: Evidence from the U.S. Vocational Rehabilitation Services Program." *Journal of Social Work in Disability and Rehabilitation*. 11(1): 33-54.

[72] Stapleton, David and William Erickson (2004). *Characteristics of Incentives: Why Do Employment Outcomes for the SSA Beneficiary Clients of VR Agencies Differ, on Average, from Those of Other Clients?* Washington DC: Cornell University Institute for Policy Research.

[73] Stapleton, David and Frank Martin (2012). "Vocational Rehabilitation on the Road to Social Security Disability: Longitudinal Statistics from Matched Administrative Data." Unpublished manuscript.

[74] Stern, Steven (1992). "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models." *Econometrica*. 60(4): 943-52.

[75] Stern, Steven (1997). "Simulation-Based Estimation." *Journal of Economic Literature*. 35(4): 2006-2039.

[76] Stern, Steven (2010). "Moments and Extrema of Estimated Marginal Effects for 'The Effects of Vocational Rehabilitation for People with Mental Illness."
http://www.people.virginia.edu/~sns5r/resint/wrkdisstf/deanstruct.html.

[77] Tremblay, T., j. Smith, H. Xie, and R. Drake (2006). "Effect of Benefits Counseling Services on Employment Outcomes for People with Psychiatric Disabilities." *Psychiatric Services*. 57(6): 816-821.

[78] U.S. Dept. of Health and Human Services, Substance Abuse and Mental Health Services Administration (1999). "Estimation Methodology for Adults With Serious Mental Illness (SMI)." *Federal Register*. 64(121): 33890-33897.

[79] U.S. Dept. of Health and Human Services, National Center for Health Statistics (2004). *National Health Interview Survey*.

[80] Uvin, J., D. Karaaslani, and G. White (2004). *Evaluation of Massachusetts' Public Vocational Rehabilitation Program: Final Report*. Massachusetts Rehabilitation Commission.

[81] Wilhelm, S. and J. Robinson (2010). *Utah State Office of Rehabilitation Economic Impact Study.* The University of Utah Center for Public Policy and Administration.

[82] Worrall, John (1978). "A Benefit-Cost Analysis of the Vocational Rehabilitation Program." *Journal of Human Resources.* 13: 285-298.