The Effects of Vocational Rehabilitation for People with Cognitive Impairments

David Dean
University of Richmond

John Pepper
University of Virginia

Robert Schmidt
University of Richmond

Steven Stern*
University of Virginia

February 22, 2013

Abstract
The public-sector Vocational Rehabilitation (VR) program is a $3 billion federal-state partnership designed to provide employment-related assistance to persons with disabilities. There is, however, relatively little-known about the long-term efficacy of VR programs. This paper utilizes unique and detailed administrative and employment data to examine both short and longer-term employment impacts for all persons diagnosed with cognitive impairments who applied for VR services in the state of Virginia in State Fiscal Year 1988 and 2000. These data provide quarterly information on VR services and employment outcomes for many years before and after the application quarter. Estimates from our model of service provision and labor market outcomes reveal that VR services generally have positive long-run labor market outcome effects that appear to substantially exceed the cost of providing services.

1 Introduction
The deinstitutionalization of persons with intellectual disabilities in the United States that started in the late 1960’s led to an increased demand for vocational services for such individuals who were in need of expanded employment opportunities (Gidugu et al., 2011). Legislation extending vocational rehabilitation (VR) assistance culminated in the landmark reauthorization of the Vocational

*We would like to thank Joe Ashley, John Phelps, Kirsten Rowe, Ann Stanfield, Vlad Mednikov, and Jim Rothrock from Virginia Department of Rehabilitation Services (DRS) for excellent help and advice and Rachel Fowley for excellent research assistance. The DRS, the Virginia Department of Medical Assistive Services, the National Institute for Disability and Rehabilitation Research, and the University of Virginia Bankard Fund for Political Economy provided generous financial support. All errors are ours.
Rehabilitation Act of 1973. VR services, previously designed for disabled veterans and persons with occupationally-related impairments, were mandated to be made available to a broader spectrum of persons with disabilities, including those with intellectual disabilities. In this paper, we provide the first evaluation of the long-term efficacy of VR service provision in the United States on the employment-related outcomes of persons with cognitive impairments.

The federal-state VR program partnership, administered by the federal Rehabilitation Services Administration (RSA), currently gives approximately $3 billion annually to state agencies to provide a wide variety of vocational rehabilitation services to individuals with a broad spectrum of disabling conditions. During the past decade, state VR agencies have closed an average of over 600,000 cases annually, with a very stable 1/9 of these being cases with a diagnosis of intellectual or developmental disability. These individuals receive a service regimen that includes human capital development ranging from pre-vocational training through supported employment. Some may even receive post-secondary education (Migliore and Butterworth, 2008b).

Yet, the last evaluation of the U.S. public-sector VR program published in an economics journal, Dean and Dolan (1991), is from over 20 years ago. More recently, several analyses have been published in the rehabilitation literature (e.g., Cimera, 2010; Bua-Iam and Bias, 2011) or produced by economic consulting firms or university research bureaus to provide state-level return on investment evaluations (e.g., Heminway and Rohani, 1999; Uvin, Karasalani, and White, 2004; Hollenbeck and Huang, 2006; Kisker et al., 2008; and Wilhelm and Robinson, 2010), but this research is seriously hampered by data and methodological limitations. The labor market outcome data is generally not of sufficient duration, either pre- or post-application, to estimate long-run VR impacts; nor do the data provide information on the intensity of the VR service provision (Loprest, 2007). Moreover, 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. Finally, the recent literature generally fails to account for the different disabilities of VR clients even though the impact of VR services is thought to differ by the type of limitation (Dean and Dolan, 1991; Baldwin, 1999; and Marcotte, Wilcox-Gok, and Redmond, 2000).

In this paper, we provide an updated and innovative evaluation of the impact of VR services on clients with cognitive impairments using unique panel

---

1 Source: Rehabilitation Services Administration, RSA-911, reported in Butterworth et al. (2011). Table 8.
2 Early economic analyses of VR efficacy include Conley (1969), Bellante (1972), Worrall, (1978) and Nowak (1983)
3 Evaluation of the efficacy of such employment services for persons with cognitive impairments began in earnest in the early 1990s with a series of randomized, controlled experiments involving transitional/supported employment training to recipients of federal Supplemental Security Income (SSI) disability payments (Prero and Thornton, 1991; Decker and Thornton; 1996). Subsequent research using small-scale demonstrations evaluate the effectiveness of training services on short-term employment outcomes (see Kregel et al., 1999; Wehman, Revell, and Kregel, 1998; Howarth et al., 2006).
data on all persons who applied for services in the state of Virginia in State Fiscal Years 1988 or 2000. For each applicant cohort, the data reveal a long panel of employment records and VR services as well as information on the client’s limitations and other characteristics at the time of application. The data on the 2000 cohort, for example, provide quarterly employment and earnings information as well as detailed information on VR services from 1995 to 2008.

Focusing attention on the important group of clients with cognitive impairments, we are able to make a number of substantive contributions to the VR evaluation literature. At the most basic level, we evaluate the short and long run labor market effects of VR services and examine the impact of specific types of services rather than just a single treatment indicator. In addition, we examine applicant cohorts where each respondent started receiving services in the same period as opposed to closure cohorts where clients terminated from the program during the same year but may have started in different periods. Finally, to address the selection problem, we formalize and estimate a structural model of endogenous service provision and labor market outcomes. We identify the parameters of this model using instrumental variables that are assumed to impact service receipt but not the latent labor market outcomes and pre-program labor market outcomes that control for differences between those who will and will not receive services.

In addition, with data on the 1988 and 2000 applicant, we are able to examine cohorts from different epochs where the services provided and clients served have notably changed. Over this period, there has been a dramatic shift in the types of training procured by VR agencies which, at one level, can be viewed as a movement away from sheltered to integrated supported employment and, at another level, as the movement from VR agencies purchasing pre-vocational and work-adjustment training in a “train and place” environment at a comprehensive rehabilitation facility to purchasing job coach training and supported employment services in an integrated work environment. Rusch and Braddock (2004) note a more than doubling in the share of supported employment participants of total day/work program participants from 1988 to 2002. At the same time, the definition of what constitutes a cognitive impairment has evolved where the 1988 cohort includes only those persons with a diagnosis of an intellectual disability while the 2000 cohort includes persons with learning disabilities, a condition that did not exist in RSA diagnostic coding in 1988. To maintain consistency in how disability is treated in the selection rule for the sample of individuals applying for services across the two periods, we restrict the sam-

\begin{itemize}
  \item[4] Dean et al. (2013a) conduct a similar analysis focused on clients with mental illnesses.
  \item[5] While these models have not been used to evaluate the VR program in the United States, several studies of the European active labor market programs for persons with disabilities have applied such methodologies (e.g., Raum and Torp, 2001; Frolich et al., 2002; Heshmati, and Lechner, 2004; and Aakvik, Heckman, and Vytlacil, 2005).
  \item[6] From a data-gathering perspective, 1988 was the first year in which a “relational” data collection system was implemented within the Virginia DRS (Vocational Rehabilitation Information System, VRIS) and this allowed, for the first time, a linkage between the purchased services by the state VR agency with the individual receiving those services.
\end{itemize}
ple analysis to persons with the same type of cognitive impairment, namely “intellectual disabilities.”

The paper proceeds as follows: Section 1.1 provides some basic background information on the number of people with cognitive impairments and the known associations between these impairments and labor market outcomes. Section 2 describes the multivariate discrete choice model for service provision choices and labor market outcomes used throughout this paper. We allow for correlation of errors among all of the equations. In Sections 3 and 4, we describe the data used in our analysis and the methodology used to estimate the parameters of the discrete choice structural model. Estimation results are presented in Section 5, and a rate-of-return analysis is presented in Section 6. The paper ends with conclusions.

1.1 Importance of the Problem

The number of individuals with cognitive impairments in the U.S. is quite large. Even a narrow classification, including only persons with “intellectual disabilities” (formerly referred to as mental retardation), results in a range from 1% to 3% of the U.S. population. Using the lower end of this range still implies some 2.5 million individuals with intellectual disabilities. Including the broader category of “developmental disability” would add in a growing number of persons diagnosed with autism spectrum disorder (ASD); certainly persons with learning disabilities or traumatic brain injury (TBI) can be considered having a cognitive impairment. Using variables available in the American Community Survey (ACS), Butterworth et al. (2011) defines a person as having a cognitive disability as “indicating that because of a physical, mental, or emotional condition lasting six months or more, s/he has difficulty learning, remembering, and concentrating” and measures 8.2 million people with a cognitive disability of working age (16 – 64) in 2009 (Butterworth et al., 2011).

People with cognitive impairments fare much worse than their non-disabled counterparts in terms of both labor force participation and subsequent employment outcomes. Butterworth et al. (2011) reports that 67.2% of the 8.2 million persons aged 16 – 64 in the U.S. with a cognitive disability were not in the labor force in 2009 (based on the ACS); the corresponding number for people with no disability is 20%. While 23.9% of individuals with cognitive impairments were employed in 2009, 71.9% of people without a disability were employed. Conditional on working in 2009, persons with cognitive impairments had mean

---

7 Defining and determining the extent of cognitive impairment is imprecise because of the broad nature of the condition and the evolving norms for its diagnosis and categorization.

8 Pierro and Thornton (1991, page 3) note that “Kiernan and Bruninks (1986) review the available evidence on the prevalence of mental retardation in the population and conclude that 1 percent of the adult population, or approximately 2 million persons, have substantial adaptive behavior limitations due to mental retardation.”

9 Lewis (2011) states that “Currently there are approximately one million American students with disabilities aged 3-21 eligible for services under the Individuals with Disabilities Education Act (IDEA) categories of intellectual disability, multiple disabilities, autism, traumatic brain injury, and developmental delay.”
annual earnings of just under $20000, while all working-age adults had mean annual earnings of $40700.

One consequence of such tenuous labor force attachment is that persons with a cognitive disability are much more likely to have income below the federal poverty levels. For 2009, while 13.1% of the U.S. working age population live in poverty, 32.9% with a cognitive impairment live in poverty (Butterworth et al., 2011). Persons with severe cognitive disabilities living in poverty are often eligible for federal disability payments in the form of Supplemental Security Income (SSI). There were 1.5 million SSI recipients of working age with cognitive impairments in 2010 receiving almost $10 billion in annual disability transfer payments. There are other costs to society and the federal and state budgets associated with the lack of employment for persons with cognitive impairments. Some are eligible for Social Security Disability Insurance, Temporary Assistance for Needy Families (TANF) cash payments, and in-kind receipt of Medicare, Medicaid, and Food Stamps. There are other costs to agencies serving persons with developmental disabilities pertaining to funding sheltered workshops and work activity centers along with adult day care and transportation programs.

2 Model

In this section, we describe the model of behavior to be estimated. It follows directly from and uses the same notation as Dean et al. (2013a). Let \( y_{ij}^* \) be the value for individual \( i \) of participating in 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

\[
\begin{align*}
y_{ij}^* &= X_i^y \beta_j + u_{ij}^y + \varepsilon_{ij}, \\
\varepsilon_{ij} &\sim \text{Logistic}
\end{align*}
\]

where \( X_i^y \) is a vector of exogenous explanatory variables, and \( u_{ij}^y \) is an error whose structure is specified below. Next, let \( z_{it}^* \) be the value to \( i \) of working at quarter \( t \), and define \( z_{it} = 1 \left( z_{it}^* > 0 \right) \). Assume that

\[
\begin{align*}
z_{it}^* &= X_{it}^z \gamma + \sum_{k=1}^K d_{ik} \sum_{j=1}^J \alpha_{jk}^y y_{ij} + u_{it}^z + \nu_{it}^z
\end{align*}
\]

where \( X_{it}^z \) 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 last quarter of service receipt and \( t \) is between \( \tau_k \) and \( \tau_{k+1} \), and \( u_{it}^z \) is an error whose structure is specified below. Next let \( w_{it} \) be the log quarterly earnings of \( i \) at \( t \), and assume that

\[
\begin{align*}
w_{it} &= X_{it}^w \delta + \sum_{k=1}^K d_{ik} \sum_{j=1}^J \alpha_{jk}^w y_{ij} + u_{it}^w + \nu_{it}^w
\end{align*}
\]

Source: SSI Annual Statistical Report 2010, Table 34 and Table 39A report 151260 recipients with autism, 271506 with developmental disabilities, and 1077484 with intellectual disabilities receiving $571.91, $609.86, and $530.28 per month, respectively, in 2010.
where variables are defined analogously to equation (2). Finally, assume that

\begin{align*}
u_{ij}^y &= \lambda_{y1}^y e_{i1} + \lambda_{y2}^y e_{i2}, \\
u_{it}^y &= \lambda_{y1}^z e_{i1} + \lambda_{y2}^z e_{i2} + \eta_{it}^y, \\
u_{it}^w &= \lambda_{w1}^y e_{i1} + \lambda_{w2}^y e_{i2} + \eta_{it}^w, \\
eta_{it} &= \rho \eta \tilde{z}_{it-1} + \tilde{\eta}_{it}, \\
eta_{it}^w &= \rho \eta \tilde{z}_{it-1} + \tilde{\eta}_{it},
\end{align*}

\begin{align*}
\begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix} &\sim iidN \left( 0, \begin{pmatrix} \sigma_{e}^2 & \rho_{e} \\ \rho_{e} & 1 \end{pmatrix} \right), \\
\begin{pmatrix} \tilde{z}_{it} \\ \tilde{w}_{it} \end{pmatrix} &\sim iidN \left( 0, I \right), \\
v_{it} &\sim iidN \left( 0, 1 \right), \\
v_{it}^w &\sim iidN \left( 0, \sigma_{w}^2 \right).
\end{align*}

We include the \((e_{i1}, e_{i2})\) to allow for two common factors affecting all dependent variables with factor loadings \(\lambda_{yk}^y, \lambda_{zk}^z, \lambda_{wk}^w\). We also allow for serial correlation and contemporaneous correlation in the labor market errors \((\eta_{it}^z, \eta_{it}^w)\). The covariance matrix of the errors implied by equation (4) is provided in Appendix (8.1).

### 3 Data

We use three main sources of data: a) the administrative records from the Virginia Department of Rehabilitative Services (DRS) for a cohort of applicants in 2000, b) the administrative records from a cohort of DRS applicants in 1988, and c) administrative records from the Virginia Employment Commission from the third quarter of 1984 to the fourth quarter of 2009 for those people in the DRS data. We also merge these records with data from the Bureau of Economic Analysis on county-specific employment patterns. Each of these is described in turn in the discussion below.

#### 3.1 DRS 2000 Applicant Cohort

##### 3.1.1 Sample Frame

We begin with the administrative records of the Virginia DRS for the 10323 individuals who applied for VR services in SFY 2000 (July 1, 1999 - June 30, 2000). Table 1 provides information about selection into our sample.\(^{11}\) The major cause of selection out of the sample is that the individual does not have a cognitive impairment. We define someone as having a cognitive impairment if either their primary or secondary diagnosis was cognitive impairment in any

\(^{11}\)Selection is performed sequentially in the order listed in Table 1.
of the service episodes observed in the data. \(^\text{12}\)  Approximately 78% of the original sample is excluded because of this restriction. We exclude another 8% of individuals who had a DRS service episode prior to 2000. We do this to avoid left censoring issues discussed in Heckman and Singer (1984). In particular, an individual participating in a subsequent service episode may be doing so because she was (endogenously) unsuccessful in the labor market or because she found the first service episode (endogenously) unusually productive. In either case, inclusion of such individuals causes estimation bias. Later, in Section 5.3, we provide some measures of the size of this bias. There are a number of other selection criteria, listed in Table 1, that have minor impacts on the final sample. After selection, we have a sample size of 1009 individuals.

### Table 1: Missing Value Analysis for 2000 Cohort

<table>
<thead>
<tr>
<th>Cause</th>
<th># Obs</th>
<th>Proportion of Total</th>
<th># Remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicants in SFY 2000</td>
<td>10323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing or Questionable SSN</td>
<td>81</td>
<td>0.008</td>
<td>10242</td>
</tr>
<tr>
<td>Died While in Program</td>
<td>65</td>
<td>0.006</td>
<td>10177</td>
</tr>
<tr>
<td>Missing Gender or Date of Birth</td>
<td>1</td>
<td>0.000</td>
<td>10176</td>
</tr>
<tr>
<td>Not Residing in Virginia</td>
<td>59</td>
<td>0.006</td>
<td>10117</td>
</tr>
<tr>
<td>No Cognitive Impairment</td>
<td>8072</td>
<td>0.782</td>
<td>2045</td>
</tr>
<tr>
<td>Missing Primary Disability</td>
<td>35</td>
<td>0.003</td>
<td>2010</td>
</tr>
<tr>
<td>Missing Secondary Disability</td>
<td>5</td>
<td>0.000</td>
<td>2005</td>
</tr>
<tr>
<td>Initial Service Spell before SFY 2000</td>
<td>822</td>
<td>0.080</td>
<td>1183</td>
</tr>
<tr>
<td>Age Younger than 17 Years</td>
<td>117</td>
<td>0.011</td>
<td>1066</td>
</tr>
<tr>
<td>Neither VR Service nor Employment Record</td>
<td>57</td>
<td>0.006</td>
<td>1009</td>
</tr>
<tr>
<td>Number Remaining in Sample</td>
<td>1009</td>
<td>0.098</td>
<td></td>
</tr>
</tbody>
</table>

\(^{12}\)In 2000 RSA coding, the equivalent of having a cognitive impairment was having a disability “cause” of mental retardation.

---

3.1.2 Service Provision

New clients to DRS are assigned a counselor who determines whether the individual is eligible for services. This step usually involves the counselor recording a diagnosis for the individual. At this point, the individual may be administratively closed because the counselor determines that the individual’s disability is insufficiently severe or that it is too severe to benefit from VR services. Also, at this point, the individual may withdraw from further consideration of service receipt. If the individual is determined to be eligible for VR services and does not withdraw, then the counselor and individual together develop an individualized plan for employment (IPE) which specifies the array of services to be provided.

The individual may receive a particular service or a combination of services. Services may include restorative medical care, training (both vocational and rehabilitative), education, job search education, and/or assistive technology services, in addition to the mandatory provision of counseling, guidance,
and placement. Services are provided a) internally by DRS personnel, b) as a “similar benefit” purchased or provided by another governmental agency or not-for-profit organization with no charge to DRS, c) as a “purchased service” through an outside vendor using DRS funds, or d) as a combination of (a), (b), and (c).

Our data come from the administrative records of DRS. These data contain the information required of each state to supply to the US Department of Education’s Rehabilitation Services Administration as the RSA-911 Case Service Report. Available data for purchased services include the beginning and ending dates of service provision, expenditures, and types of purchased services. We are not able to observe services provided in-house or through similar benefits. Based on the 1988 cohort discussed in Section 3.2, it appears that most of the in-house services were for diagnosis and evaluation.

For most of the paper, we focus exclusively on the receipt of purchased services. However, when performing a cost/benefit and rate-of-return analysis in Section 6, we impute a value of total service costs. Also, in our model, we focus on binary measures of service receipt and ignore information about either the length of service receipt or the service expenditure. We do this because we are missing some data on service expenditure and provision dates and the standard approach for evaluating labor market training and VR programs is to focus on binary indicators of service provision (see, for example, Dean and Dolan, 1991; LaLonde, 1995; Friedlander, Greenberg, and Robins, 1997; Heckman, LaLonde, and Smith, 1999; and Imbens and Wooldridge, 2009). There are 76 separate services provided by DRS, other state agencies, and 465 vendors. Since we can not estimate a set of coefficients for so many different services, nor would it be particularly useful to do so, we aggregate services, following Dean et al. (2002), into six service types listed in Table 2. Diagnosis & evaluation are provided at intake in assessing eligibility and developing an IPE. Training includes vocationally-oriented expenditures 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, eyeglasses 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.

Table 2 shows that diagnosis & evaluation and training are the two most popular purchased services. One should note that a much higher proportion of clients receive diagnosis & evaluation, but they receive it in-house. The clients who receive purchased diagnosis & evaluation may have unusually difficult cases

\footnote{We put variable names in a different font to avoid confusion.}
to diagnose and thus may be different than other clients in important unobserved ways. Thus, we have reason to be suspicious about estimates associated with diagnosis & evaluation. This is especially clear for a cohort of clients with mental illness in Dean et al. (2013a). After diagnosis & evaluation and training, restoration, maintenance, and other services are received by approximately 25% of clients, and then education is received by only 3.4% of clients. It should be noted that 5.1% of applicants with cognitive impairments in 2000 are not accepted into the program, and another 18.5% drop out after acceptance but before receiving substantive services. Thus, there are many clients who receive no services or who receive only diagnosis & evaluation.

While some clients receive no services, others receive multiple services. Table 3 provides information on the frequency of different service combinations. For example, 86 clients receive only diagnosis & evaluation (d), while 50 receive a combination of diagnosis & evaluation (d), training (t), and maintenance (m). Given the high proportion of individuals receiving more than one service type, we allow for the possibility of receipt of multiple services in our model; i.e. we have a multivariate binary choice model for service provision rather than a polychotomous discrete choice model.

### 3.1.3 Explanatory Variables

For each individual in our sample, we observe a rich set of explanatory variables listed in Table 4. Besides the usual demographic variables such as gender, race, and age, we observe some disability measures not commonly included in other data sets and some other variables particularly relevant for this population. We have dummy variables for hearing/speech disability, musculo/skeletal disability, internal disability, learning disability, and mental illness.\(^{14}\) Also, we have a measure of severity of the individual’s disability, evaluated by the counselor.

\(^{14}\)The existence of visual impairments and substance abuse problems 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.

---

Table 2: Proportion Receiving DRS Purchased Services by Type for 2000 Cohort

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td># Obs</td>
<td>1009 individuals</td>
</tr>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>0.447</td>
</tr>
<tr>
<td>Training</td>
<td>0.426</td>
</tr>
<tr>
<td>Education</td>
<td>0.034</td>
</tr>
<tr>
<td>Restoration</td>
<td>0.208</td>
</tr>
<tr>
<td>Maintenance</td>
<td>0.251</td>
</tr>
<tr>
<td>Other Service</td>
<td>0.249</td>
</tr>
</tbody>
</table>
recorded as either *not significant disability* (the reference case), *significant disability*, or *most significant disability*. The degree of disability is used by state VR agencies during periods of fiscal cutbacks to create an “order of selection” to prioritize service receipt among otherwise eligible VR clients. However, no such order of selection was in place in 2000.

Besides the usual demographic characteristics, we also observe whether the individual received special education services. Given our population of interest, about 33% received such services. A significant number of observations had missing information about education. Rather than delete them, we include a dummy variable for when education is missing. Finally, following Keith, Regier, and Rae (1991) and Ettner, Frank, and Kessler (1997), we include some potentially endogenous variables such as marital status, two transportation variables,15 and a dummy variable for receipt of government assistance.16

We are concerned about possible endogeneity of service provision. As described in much of the literature, individuals may have unobserved characteristics causing them to perform poorly in the labor market and to seek rehabilitation services. However, it might be that individuals have unobserved characteristics causing rehabilitation services to be unusually productive, thus leading to higher propensities to use the service and good labor market outcomes afterwards. We address this problem by using two instrumental variables for each of the six binary service provision variables. The first is the proportion of other clients of the individual’s counselor who were provided with the service. The second is the proportion of other clients of the individual’s VR field office who were provided with the service. Figures 1 and 2 display the distribution

---

15Raphael and Rice (2002) controls for the endogeneity of some transportation variables using a credible instrument and finds small endogeneity bias for employment. However, use of the instrument causes the effect of the transportation variable on wages to disappear.

16For most analyses, receipt of government assistance is endogenous because the rules associated with receipt depend critically on involvement in the labor market. However, for our population, most of the time, an individual can participate in the labor market to some degree without losing their benefits.

---

**Table 3: Service Combinations for 2000 Cohort**

<table>
<thead>
<tr>
<th>Combination</th>
<th>Frequency</th>
<th>Proportion of Total</th>
<th>Combination</th>
<th>Frequency</th>
<th>Proportion of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>86</td>
<td>0.106</td>
<td>dtmo</td>
<td>35</td>
<td>0.043</td>
</tr>
<tr>
<td>dt</td>
<td>84</td>
<td>0.103</td>
<td>etrm</td>
<td>33</td>
<td>0.040</td>
</tr>
<tr>
<td>t</td>
<td>61</td>
<td>0.075</td>
<td>eto</td>
<td>32</td>
<td>0.039</td>
</tr>
<tr>
<td>r</td>
<td>51</td>
<td>0.063</td>
<td>etm</td>
<td>30</td>
<td>0.037</td>
</tr>
<tr>
<td>dtm</td>
<td>50</td>
<td>0.060</td>
<td>Remaining</td>
<td>310</td>
<td>0.380</td>
</tr>
<tr>
<td>dto</td>
<td>43</td>
<td>0.053</td>
<td>Combinations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
1) d=diagnosis & evaluation, t=training, e=education, r=restoration, m=maintenance, and o=other service. Strings of letters imply receipt of each service in the string.
2) The sum of frequencies is 815; the remaining individuals received no DRS purchased services.
Table 4: Moments of Explanatory Variables for 2000 Cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>Demographic Variables</th>
<th>Disability Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Male</td>
<td>0.506</td>
<td>0.500</td>
</tr>
<tr>
<td>White</td>
<td>0.557</td>
<td>0.497</td>
</tr>
<tr>
<td>Education</td>
<td>6.770</td>
<td>5.423</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.332</td>
<td>0.471</td>
</tr>
<tr>
<td>Education Missing</td>
<td>0.033</td>
<td>0.176</td>
</tr>
<tr>
<td>Age (Quarters/100)</td>
<td>1.004</td>
<td>0.409</td>
</tr>
<tr>
<td>Married</td>
<td>0.041</td>
<td>0.197</td>
</tr>
<tr>
<td># Dependents</td>
<td>0.866</td>
<td>0.922</td>
</tr>
<tr>
<td>Transportation Available</td>
<td>0.460</td>
<td>0.498</td>
</tr>
<tr>
<td>Has Driving License</td>
<td>0.174</td>
<td>0.39</td>
</tr>
<tr>
<td>Receives Government Assistance</td>
<td>0.218</td>
<td>0.255</td>
</tr>
</tbody>
</table>

Figure 1: Marginal Distribution of Service Probabilities Disaggregated by Field Office

of service provision by field offices and by counselors, respectively. For both field offices and counselors, we observe significant variation in the proportion of clients who receive each of the six service types. For example, Figure 1 shows that, while 19% of field offices provide training to no more than 21% of their clients, 79% provide training to no less than 34% of their clients. The figure shows that training and diagnosis & evaluation are the most commonly provided services and education is the least commonly provided service with 64% of field offices providing education to no clients. More importantly, it shows a significant dispersion in service provision across field offices. Figure 2 has similar results for counselors but with even more variation. We can test the null hypothesis that the joint density of services within offices does not vary across offices using a likelihood ratio test. We reject the null hypothesis with a test statistic of 407.8 (with 240 df and normalized value of 7.66). For \( \chi^2 \) random variables with large degrees of freedom \( k \), one can normalize by subtracting

\[ \chi^2_k \]
one at a time, using a likelihood ratio test. The test statistic is 527.5 (with 288 df and a normalized value of 9.98). For counselors, the analogous test statistics are 957.3 (with 765 df and a normalized value of 40.92) and 2366.7 (with 918 df and a normalized value of 33.81). The fact that there is significant variation in the provision of services across offices and counselors makes our instrument viable.

3.2 DRS 1988 Applicant Cohort

In addition to evaluating the clients from the 2000 cohort, we also evaluate data from applicants in 1988. Over this period, there have been important changes in the classification of persons with cognitive impairments and in the provision of VR services to these persons. The 1988 cohort included only persons with intellectual disabilities, while, by 2000, the general definition of cognitive impairment had broadened to include persons with learning disabilities. To maintain consistency in selection rules across the two years, we restricted the samples based on the 1988 definition in both years.\textsuperscript{18} Between 1988 and 2000, several changes occurred for persons with cognitive impairments. The Americans with Disabilities Act of 1990 opened up greater employment opportunities (see, for example, DeLeire 2000; Acemoglu and Angrist 2001; Hotchkiss 2003; and Jolls 2004 for a discussion on the effect of the ADA on labor market outcomes for people with disabilities). Indeed, the overall civil rights movement

\begin{equation}
\chi^2 - k \sim N(0, 1)
\end{equation}

under $H_0$.

\textsuperscript{18}In section 3.2.3, we discuss how we control for learning disabilities in the 2000 cohort using a dummy explanatory variable.
for persons with disabilities led to significant movement away from provision of sheltered employment towards competitive supported employment in integrated community-based settings. Additionally, in 1997, federal Medicaid spending for supported employment was implemented, and the 1998 amendments to the Rehabilitation Act of 1973 authorized the provision of services to individuals with significant disabilities specifically with an emphasis on high-quality competitive employment which eliminated sheltered employment as a “successful rehabilitation” outcome.

### 3.2.1 Sample Frame

We begin with the administrative records of the Virginia DRS for the 11,596 individuals who applied for VR services in 1988. Table 5 provides information about selection into our sample. As was the case with the DRS 2000 Cohort, the major cause of selection out of the sample is that the individual does not have a cognitive impairment. Approximately 82% of the original sample is excluded because of this restriction versus 78% for SFY 2000. However, a larger share of the original cohort remains in the SFY 1988 final sample (16.4%) than for SFY 2000 (9.8%). There are two reasons for this difference. Most importantly, unlike the DRS 2000 Cohort, we cannot observe DRS service episodes prior to SFY 1988. So we cannot delete observations to avoid left censoring problems. (Later, in Section 5.3, we provide some evidence about the size of the bias caused by left censoring.) Secondly, other selection criteria that have minor impacts on the final sample have less of an impact on the SFY 1988 Cohort (eliminating 1.7% of the total cohort) than they did for the SFY 2000 Cohort (eliminating 4.1% of the cohort). After selection, we have a sample size of 1,907 DRS clients with cognitive impairments.

### 3.2.2 Service Provision

As with the 2000 applicant cohort, services are aggregated into 6 different groupings based on observed purchased services. For the 1988 cohort, however, we also observe similar benefits provision, which are almost entirely concentrated as diagnosis & evaluation and other services. Table 6 shows the proportion of clients receiving each service type in 1988 and 2000. Provision of diagnosis &

---

### Table 5: Missing Value Analysis for 1988 Cohort

<table>
<thead>
<tr>
<th>Cause</th>
<th># Obs Lost</th>
<th>Proportion of Total</th>
<th># Remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicants in SFY 1988</td>
<td>11,596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing or Questionable SSN</td>
<td>14</td>
<td>0.001</td>
<td>11,582</td>
</tr>
<tr>
<td>Died While in Program</td>
<td>70</td>
<td>0.006</td>
<td>11,512</td>
</tr>
<tr>
<td>Missing Gender or Date of Birth</td>
<td>10</td>
<td>0.001</td>
<td>11,502</td>
</tr>
<tr>
<td>Not Residing in Virginia</td>
<td>107</td>
<td>0.009</td>
<td>11,395</td>
</tr>
<tr>
<td>No Cognitive Impairment</td>
<td>9,488</td>
<td>0.818</td>
<td>1,907</td>
</tr>
<tr>
<td>Number Remaining in Sample</td>
<td>1,907</td>
<td>0.164</td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Proportion Receiving DRS Purchased Services by Type

<table>
<thead>
<tr>
<th>Variable</th>
<th>1988 Cohort</th>
<th>2000 Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>0.896</td>
<td>0.447</td>
</tr>
<tr>
<td>Training</td>
<td>0.519</td>
<td>0.426</td>
</tr>
<tr>
<td>Education</td>
<td>0.023</td>
<td>0.034</td>
</tr>
<tr>
<td>Restoration</td>
<td>0.202</td>
<td>0.208</td>
</tr>
<tr>
<td>Maintenance</td>
<td>0.372</td>
<td>0.251</td>
</tr>
<tr>
<td>Other Service</td>
<td>0.462</td>
<td>0.249</td>
</tr>
</tbody>
</table>

evaluation and other services are cut in half, but this is due mostly to having information in 1988 on services provided as a similar benefit. There are smaller reductions of service for training and maintenance, and these reductions are actually in purchased services. There is no change in provision of restoration, and education services increase. With respect to the composition of services within these broad aggregates, the most important change is with training where, as noted previously, we observe a shift away from work adjustment training usually provided as a prelude to sheltered employment (21.7% of training services in the 1988 cohort to 13.5% in the 2000 cohort) and toward job coach training services or supported employment (from 21.3% in 1988 to 32.1% in 2000). As was true in the 2000 cohort (Table 3), many clients receive multiple services. The most common service combinations (in order, using the same notation as in Table 3) are d, dtmo, dto, dtrmo, dtm, dt, do, dtro, dr, and o. Relative to the 2000 cohort, diagnosis & evaluation is a much more common service component in the 1988 cohort (because we have similar benefits data), and multiple service provision is more common.

3.2.3 Explanatory Variables

Table 7 provides information about the moments of explanatory variables in the 1988 cohort and how they differ from the 2000 cohort. The significant changes are in education which increases by 0.476 years of schooling, transportation availability (0.068), # dependents (0.086), the prevalence of mental illness (0.110), and the mix of disability severity. Figure 3 shows how the severity mix has substantially changed from providing services to more clients with moderate, 19The composition of services within each of the remaining aggregated service categories are relatively stable with a few exceptions. For diagnosis & evaluation, exams and visits were the dominant service (65% of diagnosis & evaluation) in 1988, and vocational evaluations (32%) and situation assessment/supported employment (39%) are dominant in 2000. For education, vocational/business school constituted 70% of services in 1988, and services were pretty evenly divided among college, secondary education, and vocational/business school in 2000. For restoration, psychological services is the dominant service in both years, and for maintenance, transportation support is the dominant service both years. Job guidance is the dominant service in both years for other services.
not significant disabilities in 1988 to almost all clients with significant or most significant disabilities in 2000.\footnote{These changes are also subject to the potential biases caused by including people with prior episodes of service in 1988. However, our intuition suggests that the bias in 1988 overstates the proportion of clients with \textit{significant} and \textit{most significant} disabilities.} This change is most likely in response to the federal RSA mandate to serve persons with more significant disabilities. In fact, starting in 2000, RSA performance standards established in the 1998 amendments to the Rehabilitation Act of 1973 required state agencies to serve certain minimum percentages of their caseloads with significant disabilities. Over the 1988 – 2000 period, there also was a national trend toward diagnosing people with a learning disability rather than with a cognitive impairment. Over this time period, in our sample, the prevalence of \textit{learning disability} doubled from 2\% to 4.6\%. While there were not enough observations with \textit{learning disability} in 1988 to use in estimation, there were in 2000.

### 3.3 VEC Data

Our next source of data is the administrative records of the Virginia Employment Commission (VEC). The VEC data provide information about individual quarterly earnings prior to, during, and after service receipt. Early economic analyses of VR efficacy (Conley, 1969; Bellante, 1972; Worrall, 1978; and Nowak, 1983) relied almost exclusively on the RSA-911 Case Service Report on case closures filed by state VR agencies. At the time, earnings were reported only at a) the time of referral to the VR program and b) following two months of employment after service provision. 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 collected from quarterly employ-
Table 8 describes the results of merging DRS and VEC data for the universe of applicants for DRS services in SFY 1988 and SFY 2000. The VEC returned longitudinal files containing quarterly employment data by cohort. Note that the SFY 1988 Cohort includes a minimum of 12 quarters of pre-application employment history and a minimum of 60 quarters post-application history. The SFY 2000 Cohort includes more pre-application quarters (16) but fewer post-application quarters (38). Despite the longer coverage period for the 1988 Cohort, the percentage of individuals with at least one quarter of “covered” employment is similar at 86.5% for 1988 and 88.6% for 2000. The remaining individuals were, for the entire interval, either a) unemployed or out of the labor force and/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).21

Throughout the analysis, the unit of time is a quarter (because we observe quarterly VEC labor market data). We want to avoid issues associated with varying lengths of time of service receipt and issues associated with how to interpret quarters when the individual is both receiving services and participating in the labor market. Thus, we assume that service provision requires only one quarter, and it occurs in the first period when the individual is observed receiving service in the data. In fact, the transitions out of service are of some interest by themselves and are discussed for a cohort of individuals with mental

---

21Dean et al. (2013a) provide evidence that indicates that there is a VEC “coverage gap” of about 12% compared to employment reported at the federal level in Social Security earnings data.
Table 8: Results from Merging VEC Employment Data with DRS Administrative Records

<table>
<thead>
<tr>
<th>Description</th>
<th>1988 Cohort</th>
<th>2000 Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Identifiers Provided to VEC</td>
<td>11596</td>
<td>10323</td>
</tr>
<tr>
<td>Application Period</td>
<td>7/87 - 6/88</td>
<td>7/99 - 6/00</td>
</tr>
<tr>
<td>Period Covered by VEC Data</td>
<td>7/84 - 6/03</td>
<td>7/95 - 12/09</td>
</tr>
<tr>
<td>Number of Quarters Covered</td>
<td>76</td>
<td>58</td>
</tr>
<tr>
<td>Individuals Returned from VEC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>10036</td>
<td>9143</td>
</tr>
<tr>
<td>Percent of Those Provided</td>
<td>86.50%</td>
<td>88.60%</td>
</tr>
</tbody>
</table>

illness in Dean et al. (2013a). For the 2000 cohort, this rule results in 16 to 19 quarters of pre-service earnings quarters and 28 to 31 post-service quarters.

The VEC data provide us with two labor market outcome variables: employment and log quarterly earnings. The employment variable is a binary measure of working in a particular quarter in the labor market, and it corresponds to $z_{it}$ associated with equation (2). The log quarterly earnings variable corresponds to $w_{it}$ in equation (3). Unfortunately, we cannot decompose quarterly earnings into wage level and hours. Thus, we directly estimate neither a wage equation nor an hours equation.

Table 9 provides information on the moments of the two labor market variables for both cohorts. The log quarterly earnings moments are conditional on employment. First, note that the large number of quarterly observations allows us to estimate precisely a rich model of labor market outcomes even after allowing for time dependence. Next, note that, in both cohorts, mean employment rates and mean earnings increase from pre-service to post-service. Also, the increases are larger in the 2000 cohort than in the 1988 cohort. This may be due to real changes in the VR program, real changes in the composition of VR applicants, and/or changes in sampling. We explore these possibilities in Section 5.3.

Erickson and Lee (2005) report an employment rate of 30.7% for adults, ages 16 - 64, with mental disabilities in 2000, and Erickson, Lee, and van Schrader (2010) report an employment rate of 28.2% for adults, ages 18 - 64, with cognitive impairments in 2008. While our post-service employment estimates are somewhat similar to national estimates, earnings results are quite different. Erickson, Lee, and van Schrader (2010) report median annual earnings of $30600 for non-institutionalized persons, aged 21 - 64, with a cognitive disability who were working full-time/full-year in 2008. Our mean log quarterly earnings estimate conditional on some employment is 7.175, which translates into annual earnings of $4 \ast \exp \{7.175\} = 5225$. Probably, the big difference between the two figures is the conditioning by Erickson, Lee, and van Schrader (2010) on working full-time/full-year; in fact, they report that only 14.0% of the same group have full-time/full-year jobs.
Figures 4 and 5 display the time series trends in the mean quarterly labor market outcomes – employment and earnings conditional on employment – for the 2000 applicant cohort. Normalizing the application quarter to equal zero, the figure illustrates how the labor market outcomes vary before and after the application quarter in SFY 2000 between applicants receiving substantial VR services – the treated group – and those that did not receive substantive services, the untreated. Prior to the application quarter, employment rates and average quarterly earnings of the treated and untreated are nearly identical. For example, one year prior to the application quarter the employment rates for both groups are about 27% and average quarterly earnings are around $1290. Thus, there is little evidence of selection into VR service receipt based on prior labor market outcomes. Shortly after the application quarter, however, the labor market outcomes of the treated and not-treated begin to diverge. In particular, Figure 4 shows that the employment rates for the treated increase relative to the rates for the untreated, leading to an employment gap of about 7% that last for 20 quarters, while Figure 5 shows that average quarterly earnings of the treated fall about $500 below the earnings of those not treated. Thus, the data reveal that VR treatment services are associated with a notable and sustained increase in employment but a drop in quarterly earnings. Although interesting descriptive statistics, the results displayed in Figures 4 and 5 should not be interpreted as measuring the impact of VR services. These models do not control for any observed covariates, do not account for selection into VR service receipt, and do not allow for heterogeneity in the types of services.

### 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 DRS data or the VEC data, the DRS data contain geographic identifiers so that we can match each DRS client with their county of residence. The Bureau of Economic Analysis (BEA) provides information on population size and
Figure 4: Employment Rates

Figure 5: Average Quarterly Earnings for the Employed
number of people employed, disaggregated by age and county (BEA, 2010). We construct measures of log employment rates using two units of geography: county and MSA/RSA level. Details are included in Dean et al. (2013a). Moments are provided in Table 10 for the 2000 cohort. The correlation between the two measures is 0.99; thus we use only the county variable in estimation.

### Table 10: Moments of Labor Market log Employment Rate Variables

<table>
<thead>
<tr>
<th>Geography</th>
<th># Obs</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>County</td>
<td>58522</td>
<td>-0.520</td>
<td>0.225</td>
<td>-1.497</td>
<td>0.096</td>
</tr>
<tr>
<td>MSA/RSA</td>
<td>58522</td>
<td>-0.515</td>
<td>0.224</td>
<td>-1.497</td>
<td>0.099</td>
</tr>
</tbody>
</table>

4 Econometric Methodology

4.1 Likelihood Function

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

\[
\theta_y = (\beta_j, \lambda_{j1}^y, \lambda_{j2}^y)_{j=1}^J, \\
\theta_z = (\gamma, \lambda_{1z}, \lambda_{2z}, \rho_\eta, \sigma_\xi^2, \rho_\zeta, [\alpha_{jz}]_j, \text{ and} \\
\theta_w = (\delta, \lambda_{1w}, \lambda_{2w}, \rho_\eta, \sigma_\omega^2, [\alpha_{jw}]_{j=1}^J).
\]

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

\[
L_i = \int L_i \left( u_i \right) dG \left( u_i \mid \Omega \right)
\]

where

\[
L_i \left( u_i \right) = L_i^y \left( u_i^y \right) \prod_{t=1}^T L_{it}^w \left( u_{it}^z, u_{it}^w \right),
\]

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

\[
L_{it}^w \left( u_{it}^z, u_{it}^w \right) = \left[ L_{it}^0 \left( u_{it}^z, u_{it}^w \right) \right]^{1-z_{it}} \left[ L_{it}^1 \left( u_{it}^z, u_{it}^w \right) \right]^{z_{it}},
\]

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

\[
L_{it}^1 \left( u_{it}^z, u_{it}^w \right) = \Phi \left( X_{it}^z \gamma + \sum_{k=1}^K d_{ik} \sum_{j=1}^J \alpha_{jk} y_{ij} + u_{it}^z \right).
\]
\[ L_{it}^z (u_{it}^z, u_{it}^w) = \frac{1}{\sigma_w} \phi \left( \frac{w_{it} - X_{it}^w \delta - \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^w y_{ij} - u_{it}^w}{\sigma_w} \right), \quad (9) \]

\[ \Phi \left( X_{it}^z \gamma + \sum_{k=1}^{K} d_{ik} \sum_{j=1}^{J} \alpha_{jk}^z y_{ij} + u_{it}^z \right), \]

and \( G(u_i \mid \Omega) \) is the joint normal density with covariance matrix \( \Omega \) described in equation (12). While, in general, it is difficult to evaluate the multivariate integral in equation (5), 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 VR service choices, \( L^y_i (u^y_i) \) in equation (6), 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, \( L^{zw}_{it} (u_{it}^z, u_{it}^w) \) in equations (7), (8), and (9), follow from the normality assumption for \( (z_{it}, w_{it}) \) and the bivariate normality assumption for \( (z_{it}, \zeta_{it}^w) \) in equation (4). 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) show that MSL estimates perform well for small and moderate numbers of draws as long as good simulation methods are used,\(^{22}\) and Geweke (1988) shows that the simulation error occurring in simulation-based estimators is of order \((1/n)\) when antithetic acceleration is used.

### 4.2 Identification

We face two types of identification issues. The first is the typical issue of identification in non-linear models. As is usual, covariation in the data between dependent variables and explanatory variables identifies many of the model parameters. For example, the \( \beta \) terms in equation (1) are identified by covariation between explanatory variables in the equation and observed VR service choices. Similar arguments apply to the \( \gamma \) terms in equation (2) and the \( \delta \) terms in equation (3). Second moment parameters such as \( \sigma^2_i \) and \( \rho_c \) in equation (4) are identified by corresponding second sample moments.

The second issue concerns identification problems caused by endogeneity of the VR service treatment choices. We address endogeneity issues in two ways. First, we control for pre-treatment labor market differences between those who do and do not receive services. (Meyer, 1995; Heckman, LaLonde, and Smith, 1999, Section 4). Second, we include two instruments for each binary service

\(^{22}\)We simulate all errors except for \( \eta \) and \( \varepsilon \) 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).
variable. The instruments for service $j$ are the propensity of an individual’s VR counselor to assign other clients to service $j$ and the propensity of an individual’s VR field office to assign other clients to service $j$. Doyle (2007), Arrighi et al. (2010), Dean et al. (2013a, 2013b, 2013c), and Clapp et al. (2010) use a similar instrument. We use a non-linear transformation of these variables, described in Appendix 8.2, to improve their performance.

In order for these variables to be valid instruments, a) they must be correlated with service receipt, b) they must not belong directly in equations (2) and (3), and c) they must be exogenous. By construction, (a) is satisfied as long as there is enough independent variation in the two instruments. Figures 1 and 2 show variation in each instrument, and the correlation between the instruments for each service is approximately 70%. There is no reason to think that (b) is a problem. Condition (c) could be a problem if counselors and/or field offices made service provision decisions based on idiosyncratic features of the local labor market. For example, a particular field office may not provide education services because there are very few local jobs that would benefit from improved education. It also could be a problem if, for example, high quality counselors both tend to choose particular services and are better than average at placing their clients in good jobs. However, despite these concerns, along with Aakvik, Heckman, and Vytlacil (2005) and Dean et al. (2013a, 2013c), this is one of the first studies to identify the impact of VR services on labor market outcomes using both a history of pre-program earnings and plausibly exogenous instrumental variables.

5 Estimation Results

In this section, we present estimation results for the 2000 cohort and compare these results to the findings from the 1988 cohort. We divide up the discussion into different parts to focus the discussion. First, we discuss estimates directly associated with the effect of purchased DRS services on labor market outcomes. Then we discuss estimates of the determinants of service choice followed by a discussion of the effect of other explanatory variables on labor market outcomes. Next, we present and interpret covariance term estimates. After presenting all of the estimates, we provide results of goodness-of-fit tests for service provision and employment and Lagrange Multiplier tests for a number of potential extensions of the model. We finish with a discussion of how estimates from the 2000 cohort differ from estimates from the 1988 cohort.

5.1 Estimates for 2000 Cohort

5.1.1 Effect of Purchased DRS Services on Labor Market Outcomes

Tables 11 and 12 report estimates of the effects of purchased DRS services on employment propensity and log quarterly earnings, respectively. Service

\footnote{Employment propensity is the latent variable in equation (2).}
effects are allowed to vary across VR service types and across the different time periods. We specify 4 different time periods: 2 or more quarters before service, the immediate quarter before VR service, the 8 quarters after service (short run), and 9 or more quarters after service (long run). Consider the effect of training in Table 11. The first column for training implies that, holding all else constant, training lowers pre-service employment propensity by 0.330. This result should be interpreted as a selection effect; i.e., individuals with unusually low employment propensity are more likely to receive training. The third and fourth columns imply that, holding all else constant, training increases employment propensity by 0.364 in the short run and 0.128 in the long run. Thus, our estimates of the short- and long-run effects of training on employment propensity are 0.364 + 0.330 = 0.694 and 0.128 + 0.330 = 0.458, respectively.24 This difference-in-difference approach is common in the literature (e.g., Meyer, 1995; Heckman, Lalone, and Smith, 1999). The second column provides an estimate of the Ashenfelter pre-program dip (Ashenfelter, 1978; Heckman, Lalone, and Smith, 1999). Employment and earnings are known to drop in the periods just before individuals apply to many labor market training programs. To account for this Ashenfelter dip (Ashenfelter, 1978; Heckman, Lalone and Smith, 1999), we explicitly allow service effects to vary in the immediate quarter before service. For the most part, these results are similar to the pre-service estimates displayed in column 1, yet are statistically insignificant.

The results in Table 11 are summarized in Figure 6. Each bar in the figure

Table 11: DRS Purchased Service Participation Effects on Employment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prior to Service Participation</th>
<th>Quarter Prior to Service Participation</th>
<th>First 2 Years After Service Participation</th>
<th>More than 2 Years After Service Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>-0.056 ** (0.018)</td>
<td>-0.033 (0.147)</td>
<td>0.366 ** (0.026)</td>
<td>0.254 ** (0.014)</td>
</tr>
<tr>
<td>Training</td>
<td>-0.330 ** (0.018)</td>
<td>-0.141 (0.155)</td>
<td>0.364 ** (0.029)</td>
<td>0.128 ** (0.014)</td>
</tr>
<tr>
<td>Education</td>
<td>0.633 ** (0.069)</td>
<td>0.693 (0.549)</td>
<td>0.755 ** (0.097)</td>
<td>1.171 ** (0.040)</td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.223 ** (0.020)</td>
<td>-0.173 (0.161)</td>
<td>-0.554 ** (0.031)</td>
<td>-0.681 ** (0.016)</td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.079 ** (0.019)</td>
<td>-0.035 (0.162)</td>
<td>-0.203 ** (0.030)</td>
<td>-0.138 ** (0.015)</td>
</tr>
<tr>
<td>Other Services</td>
<td>-0.117 ** (0.019)</td>
<td>0.200 (0.163)</td>
<td>0.266 ** (0.028)</td>
<td>0.302 ** (0.015)</td>
</tr>
</tbody>
</table>

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.

24 All F-statistics testing for the joint significance of the short-term and long-term employment effects relative to the effect prior to program participation are statistically significant with p-values less than 0.0001.
Figure 6: DRS Purchased Service Effects on Employment

corresponds to a different estimate from Table 11. The short-run effect of a service is the difference in height between the bars labeled “First 2 Years After Service Participation” and “Prior to Service Participation,” and the long-run effect is the difference between the bars labeled “More than 2 Years After Service Participation” and “Prior to Service Participation.” One can see that all services other than *restoration* and *maintenance* result in both short- and long-run increases in *employment propensity*. The variation in effects over services and over time points out the value of modelling and estimating a richer model of service provision than the usual binary service provision model in most of the literature.

Table 12 and Figure 7 provide similar estimates for the effects of each purchased service on *log quarterly earnings*. For example, the short- and long-run change in *log quarterly earnings* from *training* are $0.331 - 0.122 = 0.219$ and $0.333 - 0.046 = 0.287$, respectively. Figure 7 shows positive effects on *log quarterly earnings* for all services other than *restoration* and *maintenance*.\(^{(25)}\) 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, Heckman, and Vytlacil (2005) find no evidence of employment effects of VR services in Norway. Dean et. al. (2013a) find large positive employment and earnings effects for most service types. The one notable exception is they find negative effects for *diagnosis & evaluation* for both *employment propensity* and *log quarterly earnings* for people with mental illness and argue that they are caused by a selection effect not otherwise controlled. Here, for people with cognitive impairments, we find no such negative effect.

\(^{(25)}\)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 12: DRS Purchased Service Participation Effects on Log Quarterly Earnings

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prior to Service Participation</th>
<th>Quarter Prior to Service Participation</th>
<th>First 2 Years After Service Participation</th>
<th>More than 2 Years After Service Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>-0.157 **</td>
<td>-0.021</td>
<td>0.143 **</td>
<td>0.150 **</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.179)</td>
<td>(0.037)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Training</td>
<td>-0.331 **</td>
<td>-0.188</td>
<td>-0.122 **</td>
<td>-0.046 **</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.197)</td>
<td>(0.039)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Education</td>
<td>0.296 **</td>
<td>0.006</td>
<td>0.309 **</td>
<td>0.851 **</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.566)</td>
<td>(0.113)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.160 **</td>
<td>-0.714 **</td>
<td>-0.468 **</td>
<td>-0.401 **</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.152)</td>
<td>(0.041)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.051</td>
<td>-0.013</td>
<td>-0.302 **</td>
<td>-0.120 **</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.206)</td>
<td>(0.042)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Other Services</td>
<td>0.063 *</td>
<td>0.060</td>
<td>0.371 **</td>
<td>0.287 **</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.181)</td>
<td>(0.042)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

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.

Figure 7: DRS Purchased Service Effects on log Quarterly Earnings
as the diagnosis is more straightforward than for mental illness and such individuals with developmental disabilities have likely been extensively evaluated through special education programs prior to applying for VR services.

Tables 11 and 12 and Figures 6 and 7 provide information on the effects of purchased services on the dependent variables in equations (2) and (3). While employment propensity and log quarterly earnings were appropriate dependent variables for estimation, the variables with policy interest are employment probability and quarterly earnings, each a nonlinear function of its corresponding dependent variable. Following Dean et al. (2013a), Figure 8 uses the employment propensity effects from Table 11 and the log quarterly earnings effects from Table 12 to compute the average marginal effect of each service type on labor market outcomes. In particular, for each service \( j \) with value defined in equation (1) and each labor market outcome of policy interest, \( z_{ij} \) associated with equation (2) and \( \exp \{w_{ij}\} \) associated with equation (3), we compute

\[
\frac{1}{n} \sum_{i} [v_{ijk}(1) - v_{ijk}(0)]
\]

where

\[
v_{ijk}(y_{ij}) = \left[ \frac{1}{T_{ki}} \sum_{t \in \mathcal{A}_{ki}} v_{ijkt}(y_{ij}) \right] - \left[ \frac{1}{T_{0i}} \sum_{t \in \mathcal{A}_{0i}} v_{ijkt}(y_{ij}) \right];
\]

\( y_{ij} \) is an indicator for whether service \( j \) is received by person \( i \); \( v_{ijkt}(y_{ij}) \) is the outcome measure of interest, employment (measured in probability increments) or quarterly earnings conditional on employment (measured in $1000), for person \( i \) at time \( t \) conditional on \( y_{ij} \) (with no other service being received); \( \mathcal{A}_{ki} \) is the set of quarters observed in the data for observation \( i \) before service receipt excluding the quarter preceding service \( (k=0) \), in the short run \( (k=1) \) or the long run \( (k=2) \); and \( T_{ki} \) is the number of quarters in \( \mathcal{A}_{ki} \).

Figure 8 shows positive short- and long-run improvements in both employment probabilities and quarterly earnings for all services except restoration and maintenance. This figure reveals that both the short- and long-run mean labor market effects are estimated to be positive for diagnosis & evaluation, training, education and other services, but negative for restoration and maintenance. Figure 8 has two limitations. First, it shows only mean effects while the model implies a distribution of effects caused by interactions between nonlinearity associated with equations (2) and (3) and variation in explanatory variables in the two equations. Second, it provides information separately on short- and long-run effects but provides no information on long-term discounted benefits. Figure 9 provides information about the median discounted benefits of each service and a 95% confidence region for the benefits of each service measured in $1000. Consistent with Figure 8, all services have positive present value except restoration and maintenance. Median long-run discounted benefits are around $10000 for training, diagnosis & evaluation, and other services, and they

\[26\]

In Figure 9, we arbitrarily use a 0.95 quarterly discount factor and a 10-year time horizon. Later, in Section 6, we present information on the distribution of rates of return.

26
are $36000 for education. The figure displays significant skewness in returns; in general, for those services with positive median returns, the upper bound of the 95% confidence region is approximately 20 times further away from the median than is the lower bound; and, for those services with negative median returns the lower bound of the confidence region is approximately 20 times further away from the median than is the upper bound. The skewness is caused by the convexity of the exponential function associated with transforming $w_{it}$ in equation (3) into quarterly earnings. Note also that, for each service, all discounted benefits are on the same side of zero; this follows from the additively separable effect of services in equations (2) and (3). The wide variation in benefits points to the importance of allowing other exogenous variables to affect labor market outcomes when measuring the direct effect of the treatment.

5.1.2 Effect of Explanatory Variables on Purchased DRS Service Receipt

The model in Section 2 implies a latent service value $y_{ij}$ that depends on the explanatory variables in Table 4 and the instruments discussed below Table 4. Most of the explanatory variables from Table 4 have no statistically significant effect on service provision. Exceptions include the positive effects of receives government assistance on diagnosis & evaluation, training, and maintenance, the negative effect of musculo/skeletal disability on training, and the positive effects of age, # dependents, has driving license and musculo/skeletal disability on restoration.27

On the other hand, the instruments are very influential in explaining service provision. We allow counselor and field office effects to enter into equation (1) in two ways. First, the transformed instruments for counselor and field office

27A full set of estimates and standard errors is available from the authors.
effects described in Appendix 8.2 directly enter into equation (1). We impose
the restriction that the coefficients on counselor effects and field office effects
do not vary across services. Second, as discussed in Appendix 8.2, because
of the existence of counselors and field offices with very few clients, we include
dummy variables for those cases where counselor and/or field office effects can
not be computed. There is too much multicollinearity to include all of the
missing value variables, so we include only missing counselor effects. Table
13 displays estimates for counselor and field office effects and missing value
effects. While the field office effect is statistically insignificant, the counselor
effect is statistically significant and explains much of the variation in service
provision. We are already controlling for a large set of person-specific observed
variables, so it is unlikely that the counselor effects are capturing correlation
of characteristics with a counselor caseload. In general, the missing values
variables are statistically significant and negative, suggesting that counselors
with small caseloads are hesitant to provide services.
Table 14: Labor Market Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Employment Estimate</th>
<th>Std Err</th>
<th>Log Quarterly Earnings Estimate</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.863 **</td>
<td>0.047</td>
<td>4.748 **</td>
<td>0.080</td>
</tr>
<tr>
<td>Male</td>
<td>0.212 **</td>
<td>0.010</td>
<td>0.122 **</td>
<td>0.012</td>
</tr>
<tr>
<td>White</td>
<td>0.025 **</td>
<td>0.010</td>
<td>-0.083 **</td>
<td>0.011</td>
</tr>
<tr>
<td>Education</td>
<td>0.041 **</td>
<td>0.002</td>
<td>0.072 **</td>
<td>0.003</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.547 **</td>
<td>0.027</td>
<td>0.852 **</td>
<td>0.039</td>
</tr>
<tr>
<td>Education Missing</td>
<td>-0.710 **</td>
<td>0.041</td>
<td>-0.326 **</td>
<td>0.044</td>
</tr>
<tr>
<td>Age/100</td>
<td>1.158 **</td>
<td>0.015</td>
<td>1.045 **</td>
<td>0.019</td>
</tr>
<tr>
<td>Married</td>
<td>-0.922 **</td>
<td>0.033</td>
<td>-0.494 **</td>
<td>0.030</td>
</tr>
<tr>
<td># Dependents</td>
<td>-0.010 *</td>
<td>0.006</td>
<td>-0.090 **</td>
<td>0.007</td>
</tr>
<tr>
<td>Transportation Available</td>
<td>0.335 **</td>
<td>0.030</td>
<td>0.463 **</td>
<td>0.012</td>
</tr>
<tr>
<td>Has Driving License</td>
<td>0.612 **</td>
<td>0.013</td>
<td>0.562 **</td>
<td>0.015</td>
</tr>
<tr>
<td>Receives Govt Assistance</td>
<td>0.152 **</td>
<td>0.021</td>
<td>-0.359 **</td>
<td>0.026</td>
</tr>
<tr>
<td>Hearing/Speech Disability</td>
<td>0.081</td>
<td>0.057</td>
<td>0.304 **</td>
<td>0.048</td>
</tr>
<tr>
<td>Musculoskeletal Disability</td>
<td>0.003</td>
<td>0.020</td>
<td>-0.202 **</td>
<td>0.024</td>
</tr>
<tr>
<td>Internal Disability</td>
<td>-0.516 **</td>
<td>0.023</td>
<td>-0.764 **</td>
<td>0.025</td>
</tr>
<tr>
<td>Learning Disability</td>
<td>-0.044 **</td>
<td>0.019</td>
<td>-0.218 **</td>
<td>0.019</td>
</tr>
<tr>
<td>Mental Illness</td>
<td>-0.513 **</td>
<td>0.014</td>
<td>-0.498 **</td>
<td>0.015</td>
</tr>
<tr>
<td>Other Disability</td>
<td>-0.913 **</td>
<td>0.052</td>
<td>-1.136 **</td>
<td>0.040</td>
</tr>
<tr>
<td>Disability Significant</td>
<td>-0.300 **</td>
<td>0.034</td>
<td>-0.198 **</td>
<td>0.060</td>
</tr>
<tr>
<td>Disability Most Significant</td>
<td>-0.571 **</td>
<td>0.035</td>
<td>-0.340 **</td>
<td>0.061</td>
</tr>
<tr>
<td>Local Employment Rate</td>
<td>-0.003</td>
<td>0.103</td>
<td>-0.421 **</td>
<td>0.136</td>
</tr>
</tbody>
</table>

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

5.1.3 Effect of Explanatory Variables on Labor Market Outcomes

The model in Section 2 implies labor market outcome specifications depending on service provision (see Tables 11 and 12), the explanatory variables listed in Table 4, and the local labor market condition described in Section 3.4. Table 14 provides estimates for the explanatory variables excluding service provision. There are two estimated equations, one for employment propensity and one for log quarterly earnings. Because of the large sample sizes, almost all of the estimates are statistically significant. Some of the estimates for demographic variables have expected signs such as the positive effects of male, education, age, transportation available, and has driving license on both outcomes, and the positive effect of white on employment propensity. On the other hand, counterintuitive estimates include the negative effect of white on log quarterly earnings and the positive effect of special education on both outcomes. For disability measures, internal disability, learning disability, mental illness, and other disability have negative effects on both outcomes, and musculoskeletal disability has a negative effect on log quarterly earnings. Only hearing/speech disability has a statistically significant positive effect on any labor market outcome (log quarterly earnings). The disability severity measures work as expected: the more significant the disability, the larger the negative effect it has on both labor market outcomes. The measure of local labor market employment rate has a counterintuitive sign. In fact, in Dean et al. (2013a, 2013b, 2013c), we have consistently found no meaningful effect for local labor market conditions.
5.1.4 Covariance Terms

In general, we allow for covariation among unobservables in two ways in the model. Equation (4) specifies the existence of two latent factors affecting service provision, employment propensity, and log quarterly earnings. Table 15 presents the estimates of the factor loadings for each latent factor. Note that both factors have similar factor loading structures; i.e., with one exception, factor loadings for service provision are statistically insignificant, and the factor loadings for the two labor market outcomes have opposite signs. This suggests two factors capturing a measure of desire to work. The larger the factors, the lower the individual's reservation wage, causing employment probability to rise and quarterly earnings to fall. The fact that essentially none of the service provision factor loadings are statistically significant suggests that endogeneity of service provision may not be a large issue. Equation (4) introduces a few more covariance terms, and their estimates are displayed in Table 16. The correlation terms $\rho_n$ and $\rho_c$, allowing for serial correlation in labor market outcomes and contemporaneous correlation between the two labor market outcomes, are both statistically significant and positive. Note that the serial correlation estimate ($\rho_n = 0.977$) is similar in magnitude to estimates in the literature for non-disabled populations (e.g., Macurdy, 1982; Abowd and Card, 1989; Topel, 1991; and Meghir and Pistaferri, 2004) The positive contemporaneous correlation estimate ($\rho_c = 0.727$) suggests the existence of unobserved ability affecting employment and earnings in the same direction.

5.2 Specification Tests

5.2.1 Goodness-of-Fit Tests

We perform a number of specification tests to measure how well our model fits the data. First, we perform goodness-of-fit tests for service provision and em-
Table 16: Other Covariance Terms

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Err</th>
<th>Variable</th>
<th>Estimate</th>
<th>Std Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρη</td>
<td>0.977 **</td>
<td>0.000</td>
<td>ρζ</td>
<td>0.727 **</td>
<td>0.006</td>
</tr>
<tr>
<td>σζ</td>
<td>0.163 **</td>
<td>0.001</td>
<td>σw</td>
<td>1.096 **</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes:
1. Double-starred items are statistically significant at the 5% level.
2. Correlation terms are estimated using the transformation, $\rho = \frac{2 \exp(\xi)}{1 + \exp(\xi)} - 1$

where $\xi$ is estimated to insure that $-1 < \rho < 1$. Standard deviations are estimated using the transformation, $\sigma = \exp(\xi)$ where $\xi$ is estimated to insure that $\sigma > 0$. Standard errors for both are derived using the delta method.

Table 17: Overall $\chi^2$ Goodness-of-Fit Statistics for Service Probabilities

<table>
<thead>
<tr>
<th>Service</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Normalized Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>79.53</td>
<td>33</td>
<td>5.73</td>
</tr>
<tr>
<td>Training</td>
<td>33.51</td>
<td>36</td>
<td>-0.29</td>
</tr>
<tr>
<td>Education</td>
<td>16.14</td>
<td>14</td>
<td>0.40</td>
</tr>
<tr>
<td>Restoration</td>
<td>20.45</td>
<td>38</td>
<td>-2.01</td>
</tr>
<tr>
<td>Maintenance</td>
<td>26.65</td>
<td>35</td>
<td>-1.00</td>
</tr>
<tr>
<td>Other Service</td>
<td>25.77</td>
<td>28</td>
<td>-0.30</td>
</tr>
</tbody>
</table>

employment.\(^{28}\) Table 17 reports $\chi^2$ test statistics and normalized statistics\(^{29}\) for each service provision probability. Only the test for diagnosis & evaluation is statistically significant. Thus, overall, despite the general lack of explanatory power among most of the demographic and disability variables in Section 5.1.2, we are fitting service provision well. We performed the same test for employment probabilities disaggregated into probabilities before and after service receipt. The test statistics were $\chi^2_{31} = 926.6$ for employment probabilities before service receipt and $\chi^2_{36} = 1655.0$ for employment probabilities after service receipt. Both of these test statistics are very large, suggesting some important misspecification for employment in the model. Figure 10 plots the deviations between predicted and sample employment probabilities for the two periods. In particular, the horizontal axis has different values of $p_{it} = \Pr [z_{it} = 1 \mid X^*_it, y_i]$ from equation (2), and the vertical axis plots a kernel-weighted sample mean of $z_{it} \mid p$ where $p$ is the corresponding horizontal coordinate and the argument of the kernel function is the deviation between $p_{it}$ and $p$. Deviations between the 45° line and the other two “sample lines” at any particular predicted probabil-

\(^{28}\)For each test, we divide the unit interval into 40 equally spaced subintervals, assign observations to cells based on their predicted probability, and then compute deviations between sample and predicted probabilities for each cell. Some cells are empty and thus excluded.

\(^{29}\)Normalized statistic is $(\chi^2 - k) / \sqrt{2k} \sim N(0, 1)$ where $k$ is the degrees of freedom (DF).
Empirical Probability
Predicted Probability
Predicted and Empirical Employment Probabilities
Before Initial Service
After Initial Service
45 Degree Line

Figure 10: Predicted and Empirical Employment Probabilities

ity represent that part of employment probability that we are not predicting. Overall, we are overpredicting variation in employment probabilities across the sample.

In part, this reflects the fact that there are no interesting dynamics modeled. For example, the first-order serial correlation of employment is 0.77 before initial service and 0.84 after initial service, and the first-order serial correlation of generalized employment residuals (see Gourieroux et al., 1987) is essentially the same. Even though empirically there is significant first-order serial correlation in the data, we do not include any lagged employment term in equation (2). Thus, the model explains the variation in employment with larger coefficient estimates. Instead, in Figure 11, we plot predicted within-person average employment probabilities against empirical within-person average employment probabilities. In particular, the horizontal axis has different values of $\hat{p}_t = \sum_t p_{it}/T$, and the vertical axis plots a kernel-weighted sample mean of $\hat{z}_i \mid p$ where $p$ is the corresponding horizontal coordinate and the argument of the kernel function is the deviation between $\hat{p}_t$ and $p$. While there are still significant deviations between the predicted and empirical curves, now, at least for the after-initial-service curve, we perform significantly better in predicting the observations with high employment probabilities. In fact, the relatively poor performance in the middle probabilities is more evidence that modelling serial correlation is important. Another possibility explaining deviations is the existence of outliers, i.e., observations with very large log likelihood contributions leading to undue influence of those observations. There are some observations with large likelihood contributions for employment probabilities on the order of $-50$ (relative to a median of $-19$). However, these are likelihood contributions including approximately 55 employment probabilities, implying an average log probability of $-0.9$ with corresponding probability of $\exp \{-0.9\} = 0.4$; in fact, there are no true employment outliers. Thus, we have no complete explanation for the problematic fit for employment probabilities.
5.2.2 Lagrange Multiplier Tests

Next, we construct and perform a number of Lagrange Multiplier (LM) tests to search for important potential missing pieces of our model. In each case, we report two types of statistics. Let $\log L_i (\theta, \xi)$ be the log likelihood contribution from equation (5) where $\theta$ is the vector of parameters defined above equation (5) and $\xi$ is the vector of parameters associated with a particular hypothesis test; i.e., $H_0 : \xi = 0$ vs $H_A : \xi \neq 0$ which, in the context of the LM test, becomes $H_0 : \partial \log L_i (\theta, \xi) / \partial \xi = 0$ vs $H_A : \partial \log L_i (\theta, \xi) / \partial \xi \neq 0$. The standard LM test statistic is

$$ D^{-1} \left[ \sum_i \frac{\partial \log L_i (\theta, \xi)}{\partial \xi} \right] ' \left[ \sum_i \frac{\partial \log L_i (\theta, \xi)}{\partial \xi} \right] \sim \chi^2_k (10) $$

where $D^{-1} [\cdot]$ is the inverse covariance matrix of its argument and $k$ is the number of elements in $\xi$. However, for many of the hypotheses considered, $D^{-1} [\cdot]$ is not well-behaved because the score statistics are too collinear. Thus, we also report t-statistics,

$$ \frac{\sum_i \frac{\partial \log L_i (\theta, \xi)}{\partial \xi}}{\sqrt{\text{Var} \left[ \sum_i \frac{\partial \log L_i (\theta, \xi)}{\partial \xi} \right]}} \sim N (0, 1) (11) $$

for each individual element of $\xi$.30

In Table 14 and in our (unreported) specification of the service provision equation, we included marry and # dependents as explanatory variables but did not allow for interactions between them and male. At least for labor market

---

30It should be noted that $D^{-1} [\cdot]$ in equation (10) depends on the covariance matrix of all of the parameter estimates $(\theta, \xi)$, while the denominator in equation (11) is independent of covariances with other elements of $(\theta, \xi)$. 

---
outcomes, much of the literature (e.g., Ettner, Frank, and Kessler, 1997; Kimmel and Kniesner, 1998) suggests such an interaction. Unfortunately, for both VR service choices and employment outcomes, there wasn’t enough variation in the data with respect to marry and # dependents conditional on male = 1 to perform valid overall tests. None of the interaction terms had statistically significant individual effects for any of the dependent variables. To a large degree, this occurs for our population because the marriage rate is so much lower than in the general population of adults.

Next, we construct LM tests to determine if it is important to allow for interactions among service effects on labor market outcomes. The model naturally allows for some interactions because of the nonlinear transformations associated with equations (2) and (3). However, it might be important to allow for more general or direct interactions. We added a full set of interactions involving each pair of services to equations (2) and (3). The results of the corresponding LM test show that service interactions are not important. In fact, none of the individual score statistics associated with each potential interaction are statistically significant.\footnote{The overall $\chi^2$ test statistic could not be computed because the covariance matrix of score statistics was singular.}

One might consider a more parsimonious specification for interactions. We report the results for two such specifications. First, we include a dummy variable for each labor market outcome that is equal to one iff the individual used at least two services. Neither of the score statistics is statistically significant, and, overall, $\chi^2_3 = 2.67$ is not statistically significant. Second, we distinguish between diagnosis & evaluation versus the other five services. In particular, for each labor market outcome we include both the dummy in the first case, and we add a second dummy equal to one iff the individual used at least two services excluding diagnosis & evaluation. None of the results are statistically significant for this case either. Overall, the results suggest it is not that important to allow for service interactions.

Next, we test whether length of service provision affects labor market outcomes. This is a tentative first step in exploring the value of modeling the effects of treatment intensity, measured either as the duration over which services are provided or the expenditures spent on services, on employment outcomes. The $\chi^2_{12}$ test statistic is 39.4 which has a normalized value of 5.6. The mean score statistic and standard errors associated with this test are reported in Table 18. None of them are statistically significant, suggesting that service duration interacts with other variables already in the model.

5.3 Estimates of Change from 1988 to 2000

As noted above, there have been important changes to the mix of clients and service provisions over time. Recall, the 1998 amendments to the Rehabilitation Act of 1973 mandated that state agencies serve a more significantly disabled clientele. Also, state agencies could no longer take credit for closures into sheltered workshops, i.e. extended employment services; rather, they altered their
Table 18: Lagrange Multiplier t-Tests for Service Duration Effects on Labor Market Outcomes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Employment Probability Mean</th>
<th>Derivative</th>
<th>Std Error</th>
<th>log Quarterly Earnings Mean</th>
<th>Derivative</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis &amp; Evaluation</td>
<td>-0.816</td>
<td>0.559</td>
<td>0.071</td>
<td>0.071</td>
<td>0.232</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>-0.804</td>
<td>0.762</td>
<td>-0.130</td>
<td>-0.130</td>
<td>0.274</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>-0.134</td>
<td>0.217</td>
<td>-0.023</td>
<td>-0.023</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>Restoration</td>
<td>-0.311</td>
<td>0.511</td>
<td>0.009</td>
<td>0.009</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td>Maintenance</td>
<td>-0.356</td>
<td>0.566</td>
<td>0.013</td>
<td>0.013</td>
<td>0.217</td>
<td></td>
</tr>
<tr>
<td>Other Services</td>
<td>-0.752</td>
<td>0.548</td>
<td>-0.004</td>
<td>-0.004</td>
<td>0.209</td>
<td></td>
</tr>
</tbody>
</table>

service packages to include supported employment provided in an integrated community setting. To assess how these changes impact the evaluation of DRS services, we re-estimate the models using data from the 1988 cohort. We have 1988 cohort data analogous in structure to the 2000 cohort data that allow us to perform such an exercise. However, an important concern is that, while for the 2000 cohort data, we are able to control for left censoring (see Section 3.1.1) by excluding observations with service episodes prior to 2000, the 1988 cohort data does not provide us with the information necessary to use the same exclusion criterion. In the interest of understanding (and possibly controlling for) the bias introduced by left-censoring, we estimate the model with three samples:

a) the 1988 cohort including left-censored observations,
b) the 2000 cohort including left-censored observations,\(^{32}\) and
c) the 2000 cohort excluding left-censored observations.

The difference in estimates associated with samples (b) and (c) provides us with some information about the direction and magnitude of bias caused by left-censoring, while the difference in estimates associated with samples (a) and (b) provides us with information about the change in parameters over time.\(^{33}\)

\(^{32}\)The 2000 cohort including observations with left-censoring has 1788 observations. For the most part, it has similar moments to the cohort excluding left-censored observations. Exceptions include: mean education level decreases by 0.8, mean special education increases by 0.06, mean significant disability increases by 0.1, mean most significant disability decreases by -0.16, mean transportation available increases by 0.05, mean bus driver’s license increases by 0.06, mean diagnosis & evaluation decreases by -0.081, mean training decreases by -0.079, mean employment prior to initial service increases by 0.063, mean log quarterly earnings prior to initial service increases by 0.063, mean employment after initial service increases by 0.044, and mean log quarterly earnings after initial service decreases by -0.078. Keeping in mind that the 2000 cohort excluding left-censored observations is a subset of that including left-censored observations, some of these changes are quite large. For example, the mean reduction in training among the observations with left-censoring is -0.181 (the change in mean multiplied by 1788/(1788 - 1009) = 2.30).

\(^{33}\)Changes in the estimates for diagnosis & evaluation and other services may partially reflect differences in the way service receipt is measured in the two cohorts. Recall that similar benefits provisions are included in the service receipt indicators of the 1988 cohort but are not observed in 2000 (see Section 3.2.2). As a result, the estimates in Table 6 show that much larger fractions of the 1988 caseload are classified as receiving diagnosis & evaluation...
Figure 12: Estimated Employment Effects Across Samples

Figure 12 provides information about point estimates associated with short- and long-run service effects for each service and for each sample. Each bar represents the point estimate of the difference between either a short- or long-run effect minus the corresponding pre-service effect. For example, the estimates for the effect of training on employment for sample (a) are $-0.162$ in the pre-service period and $-0.010$ in the first two years after service; thus the estimated short-run effect is $0.162 - 0.010 = 0.152$, the height of the first bar among the training bars, which is for the short-run effect using sample (a). The pattern for short-run employment effects for training is that the estimates for samples (b) and (c) are very close, and both are significantly larger than for sample (a). This strongly suggests that, for the short-run employment effect of training, left-censoring bias is not very large, and the effectiveness of training, with respect to its effect on employment, increased significantly between 1988 and 2000. The same pattern holds true for the long-run effect of training on employment. On the other hand, for the long-run effect of education on employment, the estimates for samples (a) and (b) are close and significantly less than for (c), suggesting that, for this case, left-censoring bias explains almost all of the variation in the change in the effect. Figure 12 shows that diagnosis & evaluation increases employment in both cohorts. The effect is smaller in the 2000 cohort, and left-censoring bias hides some of the loss between the two years. For education, all effects are positive in both cohorts. In the short-run, the positive effect declined, and, in the long-run, it increases but probably only because of left-censoring bias. Restoration has negative effects on employment in both cohorts, and they seem to be declining with time. The effects for maintenance are generally small. Finally, other services has positive effects (89.6% versus 44.7%) and other services (46.2% versus 24.9%). However, it is not obvious how the change would bias estimates.
in both cohorts, and the apparent improvement over time probably is due to left-censoring bias.

Figure 13 provides analogous information for the effects of services on log quarterly earnings. For this case, diagnosis & evaluation, training, and education have uniformly positive effects. The improvement over cohorts for long-run education probably are due to left-censoring bias, while the improvement over cohorts for long-run training and the decline over cohorts for short-run education and long-run diagnosis & evaluation are real. Similar to employment effects, restoration generally has negative effects with unclear changes over cohorts. Maintenance and other service effects are difficult to interpret.

We can test to see if bias due to left censoring exists by constructing a Hausman test statistic

\[
H = n \left( \hat{\theta}_{1nc} - \hat{\theta}_{Exc} \right)' D^{-1} \left[ \sqrt{n} \left( \hat{\theta}_{1nc} - \hat{\theta}_{Exc} \right) \right] \left( \hat{\theta}_{1nc} - \hat{\theta}_{Exc} \right)
\]

where \( \hat{\theta}_{1nc} \) is the vector of parameter estimates when left-censored observations are included, \( \hat{\theta}_{Exc} \) is the vector of parameter estimates when left-censored observations are excluded, \( D^{-1} [\cdot] \) is the asymptotic covariance matrix of its argument, and \( n \) is the sample size. Under the null hypothesis that the excluded observations have the same distribution as the included observations, \( H \sim \chi^2_k \) where \( k = 193 \) is the number of common parameters in \( \theta \).\(^{35}\) The value of the test statistic is 288210 which is statistically significant at any conventional level. Thus, we can reject the null hypothesis that left-censoring bias does not exist.

To illustrate the differences in VR service efficacy across the two cohorts, it is useful to assume that left-censoring bias is stable over cohorts (but still possibly varying over specific parameters), and then use the difference in estimates between samples (b) and (c) to identify the (cohort-constant) left-censoring bias.

\(^{35}\)Note that, under the null, \( \hat{\theta}_{1nc} \) is more efficient than \( \hat{\theta}_{Exc} \), but, under the alternative, \( \hat{\theta}_{1nc} \) is inconsistent.
One could then correct the estimates from sample (a) for left-censoring and get a consistent estimate of the real change in service effectiveness over cohorts. While the necessary assumption is somewhat restrictive, it is still worthwhile to perform this exercise. The results are displayed in Figure 14. Each bar in the figure provides an estimate of the change. For example, the first bar associated with training has a height of 0.574, implying that, after controlling for left-censoring bias, the estimated change in the short-run employment effect of training is 0.574. This procedure suggests that training and other services became more effective in both the short- and long-run between 1988 and 2000. To some degree, the training estimates may be due to the reorientation of training from sheltered employment services to supported employment services, resulting in a switch from sub-minimum wage payments (which may not have been covered in VEC records) to more competitive employment with payments greater than minimum wage. Meanwhile, diagnosis & evaluation became less effective for employment but more effective for earnings, and education became less effective in the short-run but more effective in the long-run for both employment and earnings. Restoration became less effective for all cases. Overall, these results suggest that there are important differences in the efficacy of the VR services across the two cohorts.

6 Rate of Return

Figure 9 provides information about the distribution of long-run discounted benefit of each of the VR services. In this section, we add cost to the analysis and measure rates of return. As is true for Figure 9, we present information about the distribution of rates of return especially because there is wide variation in rates of return across DRS clients caused by variation in characteristics affecting labor market outcomes (Table 14). We focus this analysis on the 2000 cohort.

In general, we can think of the net benefit of service provision as the long-
run discounted benefits minus costs. We simulate the private labor market benefits to DRS clients using the structural model estimates summarized in Section 5.\textsuperscript{36} In particular, we compute the present discounted expected value of the provided services relative to receiving no services using both a 5- and 10-year post-treatment observation period for each individual who received some service. The estimated mean discounted benefits across DRS clients are $10979 with a standard deviation of $17987 using the 5-year window and $21175 with a standard deviation of $35684 using a 10-year window. Thus, the mean long-run discounted benefits are approximately twice the mean short-run benefits, reflecting the effect of adding an extra 5 years and the fact that the long-run (over 2 years) labor market effects generally are estimated to be much larger than the short-run effects and are assumed for the purpose of this analysis to last throughout the 10-year window.

One should note that private benefits might differ from social benefits in a number of ways. First, to the degree that DRS services are helpful in providing a client with a job, they may be preventing another individual from getting that job. Training programs for low skilled workers, however, have not been found to result in this type of labor market displacement (LaLonde, 1995). Second, there may be some social benefits associated with DRS service. Some, such as increased tax revenue and reduced welfare payments, have no real social benefit because any improvement for the government is a loss of equal size to the DRS client. However others, such as reduction in deadweight loss, reduced administrative cost of welfare program participation, and increased client happiness associated with having a job, have true social value but are much more difficult to measure. Thus we ignore them in this analysis and note the downward bias in estimated rates of return associated with this decision (see LaLonde, 1995).

Next, we focus on cost estimation. As was discussed in Section 3.1.2, there are three types of service provision: a) service provided internally by DRS personnel, b) service provided as a similar benefit (i.e., purchased or provided by another governmental agency or not-for-profit organization with no charge to DRS), and/or c) service purchased through an outside vendor using DRS funds. DRS keeps records specific to each client on purchased services but not on in-house services or similar benefits.\textsuperscript{37} Information about purchased service expenditures is summarized in Table 19. Mean expenditures for training ($1162) account for just over 70\% of the total average cost of purchased services. Note that the average cost for training is substantially less than the median long-run discounted benefits minus costs. We simulate the private labor market benefits to DRS clients using the structural model estimates summarized in Section 5.\textsuperscript{36} In particular, we compute the present discounted expected value of the provided services relative to receiving no services using both a 5- and 10-year post-treatment observation period for each individual who received some service. The estimated mean discounted benefits across DRS clients are $10979 with a standard deviation of $17987 using the 5-year window and $21175 with a standard deviation of $35684 using a 10-year window. Thus, the mean long-run discounted benefits are approximately twice the mean short-run benefits, reflecting the effect of adding an extra 5 years and the fact that the long-run (over 2 years) labor market effects generally are estimated to be much larger than the short-run effects and are assumed for the purpose of this analysis to last throughout the 10-year window.

One should note that private benefits might differ from social benefits in a number of ways. First, to the degree that DRS services are helpful in providing a client with a job, they may be preventing another individual from getting that job. Training programs for low skilled workers, however, have not been found to result in this type of labor market displacement (LaLonde, 1995). Second, there may be some social benefits associated with DRS service. Some, such as increased tax revenue and reduced welfare payments, have no real social benefit because any improvement for the government is a loss of equal size to the DRS client. However others, such as reduction in deadweight loss, reduced administrative cost of welfare program participation, and increased client happiness associated with having a job, have true social value but are much more difficult to measure. Thus we ignore them in this analysis and note the downward bias in estimated rates of return associated with this decision (see LaLonde, 1995).

Next, we focus on cost estimation. As was discussed in Section 3.1.2, there are three types of service provision: a) service provided internally by DRS personnel, b) service provided as a similar benefit (i.e., purchased or provided by another governmental agency or not-for-profit organization with no charge to DRS), and/or c) service purchased through an outside vendor using DRS funds. DRS keeps records specific to each client on purchased services but not on in-house services or similar benefits.\textsuperscript{37} Information about purchased service expenditures is summarized in Table 19. Mean expenditures for training ($1162) account for just over 70\% of the total average cost of purchased services. Note that the average cost for training is substantially less than the median long-run discounted benefits minus costs. We simulate the private labor market benefits to DRS clients using the structural model estimates summarized in Section 5.\textsuperscript{36} In particular, we compute the present discounted expected value of the provided services relative to receiving no services using both a 5- and 10-year post-treatment observation period for each individual who received some service. The estimated mean discounted benefits across DRS clients are $10979 with a standard deviation of $17987 using the 5-year window and $21175 with a standard deviation of $35684 using a 10-year window. Thus, the mean long-run discounted benefits are approximately twice the mean short-run benefits, reflecting the effect of adding an extra 5 years and the fact that the long-run (over 2 years) labor market effects generally are estimated to be much larger than the short-run effects and are assumed for the purpose of this analysis to last throughout the 10-year window.

One should note that private benefits might differ from social benefits in a number of ways. First, to the degree that DRS services are helpful in providing a client with a job, they may be preventing another individual from getting that job. Training programs for low skilled workers, however, have not been found to result in this type of labor market displacement (LaLonde, 1995). Second, there may be some social benefits associated with DRS service. Some, such as increased tax revenue and reduced welfare payments, have no real social benefit because any improvement for the government is a loss of equal size to the DRS client. However others, such as reduction in deadweight loss, reduced administrative cost of welfare program participation, and increased client happiness associated with having a job, have true social value but are much more difficult to measure. Thus we ignore them in this analysis and note the downward bias in estimated rates of return associated with this decision (see LaLonde, 1995).

Next, we focus on cost estimation. As was discussed in Section 3.1.2, there are three types of service provision: a) service provided internally by DRS personnel, b) service provided as a similar benefit (i.e., purchased or provided by another governmental agency or not-for-profit organization with no charge to DRS), and/or c) service purchased through an outside vendor using DRS funds. DRS keeps records specific to each client on purchased services but not on in-house services or similar benefits.\textsuperscript{37} Information about purchased service expenditures is summarized in Table 19. Mean expenditures for training ($1162) account for just over 70\% of the total average cost of purchased services. Note that the average cost for training is substantially less than the median long-

\textsuperscript{36}This simulation has a similar structure to the one used to compute marginal effects in Section 5 (see Figure 9). But here we compute the present discounted value of the actual treatments provided by DRS 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.

\textsuperscript{37}Dean et al. (2013a) find a way to impute missing in-hous services and similar benefits. They find that their inclusion in the estimation procedure has no significant effect on estimated labor market outcomes.
term discounted marginal benefit\footnote{Median benefits are being used because there are a few large outliers significantly increasing the mean.} of $9909 (see Figure 9). Overall, the mean costs of purchased services among all 1009 clients (33\% of whom receive no purchased services) equals $1632 with a standard deviation of $2663. 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 cost estimates for purchased services from Table 19 grossly underestimate total cost because they do not include a cost for DRS-provided services, similar benefits, and administrative costs. To estimate these costs, we use information on DRS spending by fiscal year as reported to the US RSA. These reports provide information on aggregate administrative costs, DRS-provided counseling, guidance, and placement service costs, purchased service cost, and size of the caseload for each fiscal year. Unfortunately, no information is provided on the costs associated with similar benefits. While there is some variation in the distribution of costs across years, in general, non-purchased service and administrative costs account for 55\% of total expenditures, reflecting an average cost per client of roughly $200 per month.

We use two methods to estimate non-purchased service costs:

1. Focus on the fact that purchased services account for 45\% of total VR costs. Since purchased service costs for our sample average $1632 per client, fixed costs are estimated to be $2000 (\approx 1632/0.45 − 1632) per client; or

2. Focus on the fact that the average costs of administration and non-purchased services is $200 per client-month. Since the average service episode length is 7 quarters, costs are estimated to be $4200 (200 \times 7 \times 3) per client.

There are some potential problems with both of these methods. First, it may be the case that service costs for individuals with cognitive impairment differ from those for individuals with other disabilities (e.g., compare our estimates with those for clients with mental illness in Dean et al., 2013a). If, for
example, cases of individuals with cognitive impairments have low average pur-
chased 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. Second, we are not allowing for any variation in cost across individuals. We could use actual expenditure data from our administrative DRS data, or we could estimate an expenditure equation using that data. In either case, we would be able to include heterogeneity in cost into our analysis. We choose not to do so because, in the model, there is no heterogeneity in service provision other than that captured by the six service types. We think it is important for the model of service expenditure to be consistent with the model of service provision.

Comparing estimated costs and benefits reveals that DRS services provided to people with cognitive impairments have substantial positive returns, especially in the longer run. Mean benefits range from $10979 for the short run to $21175 for the long run, while mean costs range from $3632 to $5832. Whether there is a net positive or negative return for any particular individual depends on the approach used to infer costs associated with non-purchased services and administration. However, even under the most conservative assumptions, the long-run mean benefit is estimated to exceed cost by a factor of 3.6.

Figure 15 provides information about the distribution of rate of return under a number of different assumptions. For each sample individual receiving some service, we compare the expected flow of benefits he would get with the service package he received relative to the flow of benefits he would get with no services. We approximate cost as

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

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 \( % \text{ with positive expenditure} \) in Table 19.

Figure 15 shows the distribution of quarterly rates of return. We choose to present results in terms of the distribution of rates of return rather than in terms of return on investment, the standard in much of the literature. We choose rate of return because

1. it is less sensitive to measurement error in cost, which is difficult to measure,\(^{39}\)

2. rate of return calculation does not require an arbitrary assumption about a discount rate while return on investment calculation does, and,

\(^{39}\)For rate of return, cost enters as a linear term, while, for return on investment, it enters as the denominator. Errors in denominators magnify estimator variance more than errors in linear components.
3. given standard choices of discount rates in the literature of 3% (e.g., Hollenbeck and Huang, 2006; Wilhelm and Robinson, 2010), return on investment calculations are quite sensitive to the length of horizon used,\(^3\) while, given the rates of return estimated in this section, results are not as sensitive to the length of horizon (see Figure 15).

We provide information for four scenarios: two with \(f = $2000\) and two with \(f = $4200\); and, for each assumption about \(f\), we consider a 10-year horizon and a 5-year horizon. 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.\(^4\) Thus, it is important to use long panels of earnings data such as ours when estimating rates of return.\(^5\) Focusing on the distribution curves associated with a 10-year horizon, one sees that 21.3% of clients with cognitive impairments have negative rates of return if \(f = $2000\) and 27.7% have negative rates of return if \(f = $4200\) (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 \(f = $4200\), the median rate of return is quite high at 4.6% quarterly (19.7% annually), and 20% of rates of return are above 12.9% quarterly (62.5% annually); if \(f = $2000\), the median rate of return is 7.7% quarterly (34.5% annually), and 20% of rates of return are above 19.2% quarterly (101.9% annually). 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.

7 Conclusions

Over the last few years, 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) have compared outcomes of a “treated” and “untreated” group,\(^6\) as we

\(^{3}\) Using a 3% discount rate, benefits 10 years out are discounted at 0.737. A stream of $1 benefits for 10 years is worth 1.86 times as much as a similar stream of benefits for 5 years.

\(^{4}\) 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.

\(^{5}\) Estimated rates of 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, Greenberg, and Robins, 1997; and LaLonde, 1995).

\(^{6}\) 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
do in Figures 4 and 5. These studies tend to find large positive returns to VR services, but they have a number of serious shortcomings including problems caused by censored data, the selection problem, and unaccounted-for heterogeneity in the caseload and in the services provided. 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). As in Dean et. al. (2013a, 2013b, 2013c), our analysis of the Virginia VR program addresses important limitations of these recent studies by evaluating a long panel of labor market outcomes before and after the provision of services; by formally accounting for the possibility that selection into the treatment is endogenous; by focusing on clients with a specific limitation, namely cognitive impairments; and by examining specific types of services rather than just a single treatment indicator. All of these contributions are found to be important for drawing inferences about the impact of VR services. A final contribution is that we explicitly address the biases arising from left-censoring by restricting our sample to first time applicants. Without this restriction, the rate of return analysis is biased. For example, note the effect of left-censoring on the estimated long-run effect of education on employment in Figure 12.

Overall, this innovative model and detailed administrative data reveal a much more nuanced picture of the VR program. On the one hand, as in these earlier evaluations, we find VR services have large positive returns: for cognitively impaired people, the mean long-run benefits of over $21000 exceed the mean costs by 5 times, a factor that is similar to the estimate reported in 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.
Wilhelm and Robinson (2010). At the same time, we find striking evidence of substantial heterogeneity in the efficacy of VR services across clients (see Figure 15), types of services (see Figure 9), and time periods (see Figure 14). Return on investment analyses ignoring these heterogeneities are nearly certain to present an incomplete and misleading picture of the efficacy of VR services.

8 Appendix

8.1 Covariance Structure

The covariance matrix of the errors \( u_i' = (u'_{i1}, u'_{i2}, \ldots, u'_{iJ}, u'_{i1T}, \ldots, u'_{iT}) \) implied by the structure in equation (4) is

\[
\Omega = \begin{pmatrix} A & B' \\ B & C + D \end{pmatrix}
\]

where

\[
A = \begin{pmatrix} \sum_k (\lambda_{1k}^y)^2 & \sum_k \lambda_{1k}^y \lambda_{2k}^y & \ldots & \sum_k \lambda_{1k}^y \lambda_{Jk}^y \\ \sum_k \lambda_{1k}^y \lambda_{2k}^y & \sum_k (\lambda_{2k}^y)^2 & \ldots & \sum_k \lambda_{2k}^y \lambda_{Jk}^y \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \lambda_{1k}^y \lambda_{Jk}^y & \sum_k \lambda_{2k}^y \lambda_{Jk}^y & \ldots & \sum_k (\lambda_{Jk}^y)^2 \end{pmatrix},
\]

\[
B = \begin{pmatrix} 1 & \rho_\zeta & \rho_\eta & \rho_\eta \rho_\zeta & \ldots & \rho_{T-1}^\zeta & \rho_{T-1}^\eta \rho_\zeta \\ \rho_\zeta & 1 & \rho_\eta \rho_\zeta & \rho_\eta & \ldots & \rho_{T-2}^\zeta \rho_\zeta & \rho_{T-1}^\eta \rho_{T-2}^\zeta \\ \rho_\eta & \rho_\eta \rho_\zeta & 1 & \rho_\zeta & \ldots & \rho_{T-2}^\eta \rho_\zeta & \rho_{T-1}^\zeta \rho_{T-2}^\eta \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\ \rho_{T-1}^\eta \rho_\zeta & \rho_{T-2}^\eta \rho_\zeta & \rho_{T-2}^\eta \rho_\zeta & \vdots & 1 & \rho_\zeta & \rho_{T-1}^\zeta \rho_{T-2}^\eta \\ \rho_{T-2}^\eta \rho_\zeta & \rho_{T-2}^\eta \rho_\zeta & \rho_{T-2}^\eta \rho_\zeta & \vdots & \rho_{T-2}^\zeta \rho_\zeta & 1 & \rho_\eta \rho_\zeta \rho_{T-1}^\eta \rho_{T-2}^\zeta \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \rho_\zeta \end{pmatrix},
\]

\[
D = \frac{\sigma_\zeta^2}{1 - \rho_\eta^2}
\]

and

\[
B = \begin{pmatrix} \sum_k \lambda_{1k}^y \lambda_{1k}^z & \sum_k \lambda_{1k}^y \lambda_{2k}^z & \ldots & \sum_k \lambda_{1k}^y \lambda_{Jk}^z \\ \sum_k \lambda_{1k}^y \lambda_{2k}^z & \sum_k \lambda_{2k}^y \lambda_{1k}^z & \ldots & \sum_k \lambda_{2k}^y \lambda_{Jk}^z \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \lambda_{1k}^y \lambda_{Jk}^z & \sum_k \lambda_{2k}^y \lambda_{Jk}^z & \ldots & \sum_k \lambda_{Jk}^y \lambda_{1k}^z \end{pmatrix}.
\]
8.2 Counselor and Field Office Effects

We use as an instrument in the service equation (equation (1)), 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 \( r_{ij} \) with

\[
\tilde{r}_{ij} = (1 - \omega_i) r_{ij} + \omega_i \bar{r}_j
\]

where \( \bar{r}_j \) is the mean value of \( r_{ij} \) across all counselors (offices), \( \omega_i = \kappa_i^{-1} \), and \( \kappa_i \) 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. These effects have a certain Bayesian flavor to them.

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. 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.

Figures 16 and 17 provide information about the distribution of the transformed counselor and office effects. One can see that there is significant variation in both. There are no unreasonable outliers despite zeroes for some services for some counselors and field offices because of the weighted average inherent in equation (13). The correlations between office and counselor variables is on the order of 0.7 for each of the six services.

\(^{44}\)The construction of instruments for this paper follows Dean et al. (2011a), and the discussion is almost exactly the same.

\(^{45}\)When a counselor (office) has only one other client, we treat it as missing also.
Figure 16: Distribution for Office Service Probabilities

Figure 17: Distribution for Counselor Service Probabilities
9 References

References


