0.1 Treatment Effects vs. Structural Parameters

The goals of the literature on structural equation estimation and on the estimation of treatment parameters are different. As first formulated by Marschak (1953), the goal of structural estimation is to solve a variety of decision problems.\(^1\)

- evaluating the effectiveness of an existing policy,
- projecting the effectiveness of a policy to

\(^1\) Recall the opening sentence of his seminal article: “Knowledge is useful if it helps us make the best decisions”. (Marschak, 1953, p.1).
different environments from the one where it was experienced, or

• forecasting the effects of a new policy, never previously experienced.

Let $\varepsilon$ denote an unobservable. In the most general form for $h$,

\begin{equation}
(3-1) \quad h = h(W, X, \varepsilon).
\end{equation}

An additively separable version of the Marshallian causal function (3-1) is

\begin{equation}
(3-2) \quad h = h(W, X) + \varepsilon.
\end{equation}

\begin{equation}
(3-1a) \quad h = h(W, X, \varepsilon, \theta)
\end{equation}

where $\theta$ is a low dimensional parameter that generates $h$. A parametric version of (3-2) is
(3-2a) \[ h = h(W, X, \theta) + \varepsilon. \]

(3-3) \[ h = \alpha'X + \beta \ln W + \varepsilon \]

Following Marschak, we distinguish 3 cases.

(1) The case where tax \( t \) has been implemented in the past and we wish to forecast the effects of the tax in the future in a population with the same distribution of \((X, \varepsilon)\) variables as prevailed when measurements of tax variation were made. (2) A second case where tax \( t \) has been implemented in the past but we wish to project the effects of the same tax to a different
population of $(X, \varepsilon)$ variables. (3) A case where the tax has never been implemented and we wish to forecast the effect of a tax either on an initial population used to estimate (1) or on a different population.

In case 1, we have data from the same population for which we wish to construct a forecast. In the randomized trial, persons face tax rate $t_j$ in regime $j$, $j = 1, \ldots, J$ assigned so that
\[ Pr(T = t_j \mid X, W, \varepsilon) = Pr(T = t_j \mid X, W). \]

In the sample from each regime we can identify

\[ (3-4) \quad E(h \mid W, X, t_j) \]

\[ = \int h(W(1 - t_j), X, \varepsilon)dG(\varepsilon \mid X, W, t_j). \]

For the entire population this function is
(3-4a) \(E(h \mid t_j) = \int h(W(1-t_j), X, \varepsilon) dG(\varepsilon, X, W \mid t_j).\)

(A-1) \((X, W, T) \perp \perp \varepsilon\)

then

\[G(\varepsilon \mid X, W, T) = G(\varepsilon) \]

\[E(h \mid W, X, t_j) = \int h(W(1-t_j), X, \varepsilon) dG(\varepsilon)\]

- **Support** \((W^*, X, \varepsilon)_{\text{target}} \subseteq \text{Support} (W, X, \varepsilon)_{\text{historical}}\)
0.2 Two Different Cases For Social Experiments

“Treatment” is a tax policy - say a proportional tax on wages.

\[(A-5) \quad (T \perp \perp \varepsilon) \parallel (W, X)\]

\[\Pr(T = t \mid W, X, \varepsilon) = \Pr(T = t \mid W, X)\]

\[E(h \mid t, W, X) = \int h(t, W, X, \varepsilon) dG(\varepsilon \mid t, W, X)\]

\[= \int h(t, W, X, \varepsilon) dG(\varepsilon \mid W, X)\]

\[E(h \mid t', W, X) = \int h(t', W, X, \varepsilon) dG(\varepsilon \mid W, X).\]
\[ E(h \mid t, W, X) - E(h \mid t', W, X) = \]
\[ \int [h(t, W, X, \varepsilon) - h(t', W, X, \varepsilon)]dG(\varepsilon \mid W, X). \]
Population average treatment effect for taxes 

$E_{F_c}(h \mid t) - E_{F_c}(h \mid t') =$

$$\int [E(h \mid t, W, X) - E(h \mid t', W, X)] dF_c(W, X)$$
Additive separability of $h$ in $\varepsilon$ facilitates task.

If

$$h = h(W, t, X) + \varepsilon,$$

$$E(h \mid W, X, t) - E(h \mid W, X, t')$$

$$= h(W, X, t) - h(W, X, t').$$

$$h = \alpha_0 + \alpha_1 \ln(W(1 - t)) + \alpha_2 X + \varepsilon =$$

$$\alpha_0 + \alpha_1 \ln W + \alpha_1 \ln(1 - t) + \alpha_2 X + \varepsilon.$$

$$E(h \mid W, t, X) - E(h \mid W, t', X)$$

$$= \alpha_1[\ln(1 - t) - \ln(1 - t')].$$

This is a general point about the data produced from social experiments. Social experiments only identify treatment terms and terms that
interact with treatment. Main effects for 

\((W, X)\) are not identified.
Consider the additively separable case

\[ h(W, X, t, \varepsilon) = h(W, X, t) + \varepsilon. \]

Under assumption (A-5) for randomization, and full compliance, we can recover

\[ h(W, X, t) - h(W, X, t') \]

for various treatment (tax) combinations, \( t \neq t' \).

However, decomposing \( h(W, X, t) \) into a main effect term \( \varphi(W, X) \) and an interaction term plus main effect for treatment term \( \eta(W, X, t) \) we may write

\[ h(W, X, t) = \varphi(W, X) + \eta(W, X, t). \]

\( \varphi(W, X) \) differences out in all con-
trasts. Only differences in $\eta(W, X, t)$ can be identified.
Randomization identifies the treatment effect not by creating exogeneity between the “right hand” variables and the error term and identifying the Marshallian causal parameters, but rather by balancing the bias. Thus, as a consequence of (A-5)
\[ E(\varepsilon \mid t', W, A) = E(\varepsilon \mid t, W, A). \]

\[ \varepsilon \perp \perp (W, X). \]
0.3 Two Different Cases For Social Experiments

Using the labor supply example of Section 2.2, we now demonstrate the contrasting nature of the two cases for social experiments. We also present a form of experimentation that identifies the marginal effect of policy changes (or other variables) on outcomes. The first, and historically older, case in economics seeks to use randomization to identify Marshallian causal functions and structural parameters.
The goal of this type of analysis is to form counterfactuals for policies never tried or to project the effects of policies experienced in one environment to new environments.

The second, and more recent, case seeks to use social experiments to evaluate the effectiveness of various “treatments” in place for various the treatment parameters defined in Part I, most often, treatment on the treated is the parameter of interest in these evaluations.

In the following example, the “treatment”
is a tax policy - say a proportional tax on wages.\textsuperscript{2} Thus the goal of experiments under the second case for social experiments is to determine how labor supply responds to taxes \( t \) in an experimentally determined population. No explicit attention is given to forecasting the effects of the tax on different populations or in different economic environments for the

\textsuperscript{2} Historically, randomization was first used in economics to vary wage and income parameters facing individuals in order to estimate wage and income effects in labor supply to examine the implications of negative income taxes on labor supply. Part of the goal of randomization was to produce variation in wages and incomes to determine estimates of income and substitution effects. See Cain and Watts (1973). Ashenfelter (1983) shows how estimates of income and substitution effects can be used to estimate the impact of negative income taxes on labor supply.
same population. The labor supply equation is \( h = h(t, W, X, \varepsilon) \), where \( X \) may include unearned income and asset income so that labor supply depends on both wage and unearned income. Assume that taxes \( T \) are assigned to persons so that

\[(A-5) \quad (T \perp \varepsilon) \parallel (W, X).\]

Thus \( \Pr(T = t \mid W, X, \varepsilon) = \Pr(T = t \mid W, X) \). Assuming full compliance with the assignment, we may compute the labor supply given \( t \) ("treatment" or taxes) as,
\[ E(h \mid t, W, X) = \int h(t, W, X, \varepsilon) dG(\varepsilon \mid t, W, X) = \int h(t, W, X, \varepsilon) dG(\varepsilon \mid W, X) \]

where the second equality follows from (A-5).

Similarly, for the same fixed population, but for tax rate \( t' \),

\[ E(h \mid t', W, X) = \int h(t', W, X, \varepsilon) dG(\varepsilon \mid W, X). \]

The treatment effect of taxation \( t \) relative to a base state \( t' \) on the same fixed population is

\[ E(h \mid t, W, X) - E(h \mid t', W, X) = \]

20
\[ \int [h(t, W, X, \varepsilon) - h(t', W, X, \varepsilon)] dG(\varepsilon \mid W, X). \]

We may remove the conditioning on \((W, X)\) by integrating out \((W, X)\) using a common distribution. Let \(F_c(W, X)\) denote the common distribution. It could be any benchmark, or one selected to match the features of a particular target population.\(^3\) Then the population average treatment effect for taxes \((t, t')\) is

\(^3\) There may be different distributions of \(W, X\) given \(t\) by virtue of the assignment rule producing (A-5). If assignment to treatment is the same for all \((W, X)\) (so \(T \perp \varepsilon\)), then a common distribution of \((W, X)\) is produced by randomization in all treatment categories.
\[
E_{F_c}(h \mid t) - E_{F_c}(h \mid t') = \int [E(h \mid t, W, X) - E(h \mid t', W, X)]dF_c(W, X),
\]
for all, \(t, t', t \neq t'\). These treatment effects combine structural (and causal parameters) in an economically uninterpretable form. Yet at the same time, they answer the specific question of how labor supply responds to taxes \(t\) and \(t'\) in the populations over which randomization is conducted.

Applying the results of the experiment to a new population, or forecasting the effects
of tax rates not previously experienced, requires the same types of adjustments described in Section (I.3). It is necessary to decompose $E(h \mid t, W, X)$ into $h(\cdot)$ and $G(\cdot)$ components and to determine these functions over the target supports for the distributions for the target population and the target tax rates. Structural assumptions must be invoked to extrapolate to the target population.

Additive separability of $h$ in $\varepsilon$ facilitates this task. Thus if
\[ h = h(W, t, X) + \varepsilon, \]

under assumption (A-5)

\[
E(h \mid W, X, t) - E(h \mid W, X, t') = h(W, X, t) - h(W, X, t').
\]

Then the treatment effect is the difference between two Marshallian causal functions.

With additional structure imposed, it is possible to move from treatment effects to combinations of explicit structural parameters that are determined by interactions between \( T \) and \((X, W)\) and the main effects of \( T \).

Thus suppose that we further specialize the
Marshallian causal functions so that there are only main effects in $t$:

$$
\begin{align*}
  h &= \alpha_0 + \alpha_1 \ln(W(1 - t)) + \alpha_2'X + \varepsilon = \\
  &= \alpha_0 + \alpha_1 \ln W + \alpha_1 \ln(1 - t) + \alpha_2'X + \varepsilon.
\end{align*}
$$

Recall that in this case, the derivatives of the Marshallian parameters are the slopes and the structural parameters. Under (A-5), equation (3-6) specializes to

$$
E(h \mid W, t, X) - E(h \mid W, t', X) = \alpha_1[^\frac{}{}][\ln(1 - t) - \ln(1 - t')]\]
$$

and $\alpha_1$ is identifiable from the treatment effects, [just divide both sides by the expression in brackets if $t \neq t'$]. More generally, under
additive separability and (A-5) we can identify the combinations of structural parameters represented in (3-6). Randomization governed by (A-5) does not identify $\alpha_2$. In general, $\varepsilon$ and $W$ are stochastically dependent, and the variation induced in $T$ by virtue of a randomization that implements (A-5) does not make $W$ or $X$ exogenous (independent of $\varepsilon$). The only reason why the coefficient on the wage term is identified in this example is that it is the same as the coefficient on taxes.
This is a general point about the data produced from social experiments. Social experiments only identify treatment terms and terms that interact with treatment. Main effects for \((W, X)\) are not identified. Thus consider the additively separable case \(h(W, X, t, \varepsilon) = h(W, X, t) + \varepsilon\).

Under assumption (A-5) for randomization, and full compliance, we can recover \(h(W, X, t) - h(W, X, t')\) for various treatment (tax) combinations, \(t \neq t'\). However, de-
composing \( h(W, X, t) \) into a main effect term \( \varphi(W, X) \) and an interaction term plus main effect for treatment term \( \eta(W, X, t) \) we may write

\[
h(W, X, t) = \varphi(W, X) + \eta(W, X, t).
\]

\( \varphi(W, X) \) differences out in all contrasts. Only differences in \( \eta(W, X, t) \) can be identified.

Randomization identifies the treatment effect not by creating exogeneity between the “right hand” variables and the error term and identifying the Marshallian causal parameters, but rather by balancing the bias. Thus, as a
consequence of (A-5)
\[ E(\varepsilon \mid t', W, A) = E(\varepsilon \mid t, W, A). \]

The “control functions” (or conditional bias terms) are equated across treatment groups as a consequence of randomization and can be differenced out across treatments. This is a feature of randomization shared by matching, nonparametric instrumental variables and an entire class of “control function” methods discussed below in Section 8 including panel data models. $W$ and $X$ are not exogenous, randomization of $t$ does
not make them exogenous. Exogeneity of the conditioning variables is not required to construct the treatment effect that compares the mean outcomes under the two treatments. However, as discussed in Section 2.2., exogeneity becomes an important issue if we seek to apply the results from one experiment to another environment, or if we seek to predict the effects of tax rates not previously experienced on labor supply.

Thus if we seek to project the findings
from one experiment to a new population
with the same distribution of $\varepsilon$ but different
distributions of $(W, X)$, the task is greatly
simplified by assuming
$$\varepsilon \perp \perp (W, X).$$

Then it is no longer necessary to determine the
distribution of $\varepsilon$ given $W, X$ ($G(\varepsilon \mid W, X)$)
in the target population. It is still necessary
to determine $h(W, X, t, \varepsilon)$ over the support
of the target population which remains a
formidable task.