Introduction to Econometrics II: Recitation 10[*]

Seung-hun Lee[†]

April 15th, 2020

1 Semiparametric Regresion

Semiparametrics can be thought of as a middle ground between nonparametric and parametric regression. Suppose that we partition the covariates into two unoverlapping spaces - X and W. Also assume that $E(\epsilon|X, W) = 0$. One example of a semiparametric regression is a **partially linear regression** which has the following form

$$Y_i = X_i\beta + g(W_i) + \epsilon_i$$

where $\beta \in \dim(X_i)$ represents a coefficient for the linearly regressed terms and $g(\cdot)$ is a nonparametric portion of the regression.

To estimate β, we use the fact that

$$\begin{aligned} E[Y_i|W_i] &= E[X_i|W_i]\beta + E[g(W_i)|W_i] + E[\epsilon_i|W_i] \\ &= E[X_i|W_i]\beta + g(W_i) + E[\epsilon_i|W_i] \\ &= E[X_i|W_i]\beta + g(W_i) + E[E[\epsilon_i|X_i, W_i]|W_i] = E[X_i|W_i]\beta + g(W_i) \end{aligned}$$

Given this, we can write

$$Y_i - E[Y_i|W_i] = \{X_i - E[X_i|W_i]\}\beta + \epsilon_i$$

Then we follow this procedure

[*] This is based on the lecture notes of Professors Jushan Bai and Bernard Salanie. I was also greatly helped by previous recitation notes from Paul Sungwook Koh and Dong Woo Hahm. The remaining errors are mine.

[†] Contact me at sl4436@columbia.edu if you spot any errors or have suggestions on improving this note.

1. Nonparametrically estimate $E[X_i|W_i]$ and $E[Y_i|W_i]$. Then define $\tilde{X}_i = X_i - \hat{E}[X_i|W_i]$ and $\tilde{Y}_i = Y_i - \hat{E}[Y_i|W_i]$, where $\hat{E}$ are nonparametric estimators

2. Regress $\tilde{Y}$ onto $\tilde{X}$ to get an estimate of β

3. We can estimate $g(\cdot)$ by nonparametrically regressing $Y_i - X_i\hat{\beta}$ onto W_i

As far as approximation is concerned, β follows the properties of parametric estimators. That is, it converges at rate $n^{-1/2}$ regardless of the dimensions of X_i, W_i. Thus there is no curse of dimensionality here. Unfortunately, estimating $g(\cdot)$ follows the same properties as nonparametric estimators. It has a slower convergence rate, which becomes even slower with more dimensions of W_i.

There is another caveat to this method. Identification of β requires an exclusion restriction in the sense that none of the components in X_i is perfectly predictable by W_i components. If otherwise, $X_i = E[X_i|W_i]$. Thus, β cannot be identified. This would effectively rule out including a constant in the X_i part of the regression. If we do have $X_i \neq E[X_i|W_i]$, we have

$$\begin{aligned} &(X_i - E[X_i|W_i])'(Y_i - E[Y_i|W_i]) = (X_i - E[X_i|W_i])'[(X_i - E[X_i|W_i])\beta + \epsilon_i] \\ \implies &E\{(X_i - E[X_i|W_i])'(X_i - E[X_i|W_i])\}\beta = E\{(X_i - E[X_i|W_i])'(Y_i - E[Y_i|W_i])\} \end{aligned}$$

The value of β that solves the above equation is the estimator we are looking for.

Example 1.1 (Horowitz, Lee (2002))**.** *The paper reviews several semiparametric methods for estimating conditional mean functions and show that semiparametrics allow more flexibility than parametric modeling and more precision than nonparametric models. For fun (at least for a baseball nerd like me), this paper tests this idea on a data of salaries, runs, tenure of baseball players in 1987.*

Example 1.2 (Ucal et al (2010))**.** *This paper analyzes whether and to what extent the inflow of FDI is affected before and after the occurence of a financial crisis in developing countries using generalized partial linear models. The results indicate that FDI inflows decrease in the years after a financial crisis and an upturn in FDI inflows the year before a financial crisis hit the country.*

2 Introduction to Treatment Effects

2.1 Causality and Treatment Assignment

Now, our interest is in finding out whether some X variable **causes** Y, not just whether they are correlated. Generally, correlation is not enough to argue causation. Suppose that you find that $cov(X, Y) \neq 0$. This can happen because

- X do cause Y, which is good for us. But..
- Y could also cause X. So there is a reverse causality bias here
- Z mutually affects X and Y. This is an omitted variable bias and leads to nonzero correlation even if X and Y has no connection whatsoever.

In any study examining program evaluation or quasi-experimental setting, the key issue is the assignment of the treatment, which is considered binary for the moment. There are three possible cases, which are

- **Random Assignment**: This would be a case when we can guarantee that the assignment to the treatment arms (treatment and control) are determined by chance - like flipping a coin or random draw of numbers. This rarely happens in social science, as people can voluntarily opt-in and opt-out of treatment.
- **Selection on Observables**: The treatment assignment is effectively random once we condition on some observable covariates. This is also known as ignorability or unconfoundedness assumptions.
- **Selection on Unobservables**: The assignment depends fundamentally on unobservables, or in other words, we cannot break down the dependence structure of assignment using observed variables. For instance, even in a clinical trials that involve voluntary participation, it is very likely that (lack of) risk averseness determines participation. How do we control for risk averseness?

2.2 Theoretical Framework for Treatment Effect

We will consider a binary treatment variable - whether individual i received a treatment or not (in the treatment group or control group). Define a variable D_i s.t.

$$D_i = \begin{cases} 1 & \text{If treated} \\ 0 & \text{If not treated} \end{cases}$$

In the above notation, i indexes the unit of the treatment - an individual, household, county, firm, and etc. For each unit i, there are two possible outcomes. One is the outcome without treatment, $Y_i(0)$. The other is the outcome with treatment, $Y_i(1)$. This can be seen as a counterfactual framework in the sense that if we get $Y_i(1)$ for unit i, we cannot get $Y_i(0)$ and vice versa. We always have a missing data problem in this regard. A useful way of writing this is

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

The left hand side is an outcome for unit i, treated or untreated. *This is observed for every i.* The right hand side is meant to capture that the observed outcome for individual i is either *one* of $Y_i(1)$ or $Y_i(0)$. This framework is sometimes called **potential outcome framework** and will come in handy when we analyze various treatment effect results.

We are interested in how an outcome for unit i changes between treatment and control. In other words,

$$TE_i = Y_i(1) - Y_i(0)$$

Another parameter of interest could be the treatment effect averaged over those who share the common covariate value, which is

$$TE(x) = E(TE_i | X_i = x)$$

We could also be interested in the treatment effect averaged over the population. This is the average treatment effect, defined as

$$ATE = E(TE_i) = E(Y_i(1) - Y_i(0))$$

Of course there are more, most of which we will learn over the course of this sequence. There are average treatment effect on the treated, ATE for the untreated, etc.

Because of the fundamental problem of a missing data, we are forced to make an assumption about the things we cannot observe. Take the ATE for example. We can write

$$\begin{aligned}E[Y_i(1) - Y_i(0)] &= E[Y_i(1)] - E[Y_i(0)]\\ &= \{\Pr(D_i = 1)E[Y_i(1)|D_i = 1] + (1 - \Pr(D_i = 1))E[Y_i(1)|D_i = 0]\}\\ &- \{\Pr(D_i = 1)E[Y_i(0)|D_i = 1] + (1 - \Pr(D_i = 1))E[Y_i(0)|D_i = 0]\}\end{aligned}$$

Here is the problem: We can get what $E[Y_i(1)|D_i = 1]$ and $E[Y_i(0)|D_i = 0]$ are from the data. This is because [1]

$$\begin{aligned}E[Y_i|D_i = 1] &= E[1 \cdot Y_i(1) - (1 - 1) \cdot Y_i(0)|D_i = 1] = E[Y_i(1)|D_i = 1]\\ E[Y_i|D_i = 0] &= E[0 \cdot Y_i(1) - (1 - 0) \cdot Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 0] \qquad \text{(TE)}\end{aligned}$$

So we can get $E[Y_i(1)|D_i = 1]$ and $E[Y_i(0)|D_i = 0]$ from the data - take the expected value of observed Y_i conditional on $D_i = 1$ and 0. We cannot do the same for $E[Y_i(1)|D_i = 0]$ and $E[Y_i(1)|D_i = 0]$. This forces us to make an assumption about what the two unobserved values could be.

Before going on, it would be useful to characterize the counterfactual outcome in a more econometrics-friendly context. Define the counterfactual outcomes as

$$Y_i(D_i = d) = \mu(X_i, d) + \epsilon_i(d)$$

where d can take either $0, 1$. What we want to learn involves understanding the joint distribution of the variables $Y_i(0)$ and $Y_i(1)$. Take the treatment effect at $X_i = x$, written as

$$TE(x) = \mu(x, 1) - \mu(x.0)$$

This is the effect of the treatment on the outcome Y_i for $X_i = x$. The interpretation is that if we shift everyone with $X_i = x$ from the control group to the treatment group, the average outcome increases by $TE(x)$. We can also calculate ATE as

$$\begin{aligned}ATE = E[Y_i(1) - Y_i(0)] &= E[\mu(X_i, 1) - \mu(X_i, 0) + \epsilon_i(1) - \epsilon_i(0)]\\ &= E\left[E[\mu(X_i, 1) - \mu(X_i, 0) + \epsilon_i(1) - \epsilon_i(0)|X_i]\right] = E\left[E[\mu(X_i, 1) - \mu(X_i, 0)|X_i]\right]\\ &= E[\mu(X_i, 1) - \mu(X_i, 0)] = E[TE(X_i)]\end{aligned}$$

[1] You might have noticed that I rarely tag equations in my recitation notes. So when I do, it means I am going back to this very often.

2.3 Randomized Assignment

We are back to having to assume what $E[Y_i(1)|D_i = 0]$ and $E[Y_i(1)|D_i = 0]$ are. A **random assignment** assumes that the outcome is independent of the treatment status. More formally

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \tag{RA}$$

This implies that

$$E[Y_i(1)] = E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0]$$

and similarly for $E[Y_i(0)]$. If we go back to the (TE) equation, the problem was that we were unable to make any statement about $E[Y_i(1)|D_i = 0]$ and $E[Y_i(1)|D_i = 0]$. Now what we are doing is to equate

$$\begin{aligned} E[Y_i|D_i = 1] = E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0] \\ E[Y_i|D_i = 0] = E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1] \end{aligned}$$

and effectively make the statements about the previously unobserved values. This allows us to rewrite the ATE as

$$\begin{aligned} E[Y_i(1) - Y_i(0)] &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \ (\because \text{RA}) \\ &= E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \ (\because \text{TE}) \end{aligned}$$

Therefore, we can estimate ATE by mapping $E[Y_i(1)]$ to $E[Y_i|D_i = 1]$, $E[Y_i(0)]$ to $E[Y_i|D_i = 0]$.

2.4 Conditional Independence Assumption

In this setup, we are assuming that conditional on X_i, the outcomes and D_i are independent. Formally, we can write

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i|X_i \tag{CIA}$$

We can alternatively conceptualize this framework in the following way: Define D_i as

$$D_i = 1(u_i < p(X_i))$$

where $u_i \equiv U[0, 1]$ determines selection and $p(X_i)$ can be interpreted as a propensity score - how much are you likely to be treated given that you have covariates specified as some X_i.

This framework, along with the (CIA), allows us to say the following:

$$
\begin{aligned}
E[Y_i(1)|X_i = x] &= E[Y_i(1)|D_i = 1, X_i = x] = E[Y_i(1)|D_i = 0, X_i = x] \ (\because \text{CIA}) \\
&\implies E[\mu(x,1) + \epsilon_i(1)|D_i = 1, X_i = x] = E[\mu(x,1) + \epsilon_i(1)|D_i = 0, X_i = x] \\
&\implies E[\mu(x,1)|D_i = 1, X_i = x] - E[\epsilon_i(1)|D_i = 1, X_i = x] \\
&= E[\mu(x,1)|D_i = 0, X_i = x] + E[\epsilon_i(1)|D_i = 0, X_i = x] \\
&\implies E[\epsilon_i(1)|D_i = 1, X_i = x] = E[\epsilon_i(1)|D_i = 0, X_i = x] \ (\because \mu(x,1) \text{ is same for both cases}) \\
&\implies E[\epsilon_i(1)|u_i < p(x), X_i = x] = E[\epsilon_i(1)|u_i > p(x), X_i = x] \\
&\implies (\epsilon_i(1), \epsilon_i(0)) \perp\!\!\!\perp u_i|X_i \qquad \text{(CIA2)}
\end{aligned}
$$

Thus, conditional on observed values for the covariate X_i, the unobserved elements of the selection rule u_i are independent of the unobserved elements of the error process $(\epsilon_i(1), \epsilon_i(0))$.

The caveat, however, is that $p(x) \in (0,1)$. If $p(x) = 1$, then for every possible u_i, $D_i = 1$ - everyone gets treated. Therefore, there is no meaningful statement we can make about the $D_i = 0$ case. We can say similarly for $p(x) = 0$ case. In some textbooks, this is also known as **overlap** assumption - each unit in the defined population should have some chance of being treated and some chance of being in the control group. [2]

Under CIA, we can derive the treatment effect as follows

$$
\begin{aligned}
E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x] &= E[\mu(x,1) + \epsilon_i(1)|D_i = 1, X_i = x] - E[\mu(x,0) + \epsilon_i(0)|D_i = 0, X_i = x] \\
&= E[\mu(x,1)|D_i = 1, X_i = x] + E[\epsilon_i(1)|D_i = 1, X_i = x] \\
&- E[\mu(x,0)|D_i = 0, X_i = x] - E[\epsilon_i(0)|D_i = 0, X_i = x] \\
&= \mu(x,1) + E[\epsilon_i(1)|X_i = x] - \mu(x,0) - E[\epsilon_i(0)|X_i = x] \ (\because \text{CIA2}) \\
&= \mu(x,1) - \mu(x,0)
\end{aligned}
$$

Note that

$$
\begin{aligned}
E[Y_i(1) - Y_i(0)|X_i = x] &= E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x] \\
&= (\Pr(D = 1|X_i = x) \cdot E[Y_i(1)|D_i = 1, X_i = x] + (1 - \Pr(D = 1|X_i = x))E[Y_i(1)|D_i = 0, X_i = x]) \\
&- (\Pr(D = 1|X_i = x) \cdot E[Y_i(0)|D_i = 1, X_i = x] + (1 - \Pr(D = 1|X_i = x))E[Y_i(0)|D_i = 0, X_i = x]) \\
&= E[Y_i(1)|D_i = 1, X_i = x] - E[Y_i(0)|D_i = 0, X_i = x] \ (\because \text{CIA}) \\
&= E[Y_i|D_i = 1, X_i = x] - E[Y_i|D_i = 0, X_i = x]
\end{aligned}
$$

Thus, under (CIA) we can back out the ATE for $X_i = x$ using the observables. This assumption allows us to map $E[Y_i(1)|X_i = x]$ to $E[Y_i|D_i = 1, X_i = x]$ and $E[Y_i(0)|X_i = x]$ to $E[Y_i|D_i = 0, X_i = x]$, which is impossible without it.

[2] If both (CIA) and overlap is satisfied, we say that have **strong ignorability** (Rosenbaum, Rubin 1983).

This is called a **regression adjustment** method. The treatment effect for $X_i = x$ using a regression adjustment can be obtained by utilizing the fact that

$$\mu(x,1) = E[Y_i|D_i = 1, X_i = x],\ \mu(x,0) = E[Y_i|D_i = 0, X_i = x]$$

and regressing on the subsample of each treatment arm to get $\hat{\mu}(x,1)$ and $\hat{\mu}(x,0)$. Thus

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}(x,1) - \hat{\mu}(x,0))$$

2.5 Inverse Probability Weighting

We can obtain the average treatment effect using a different approach. This is an **inverse probability weighting**. The steps are as follows

1. Estimate the propensity score $p(X_i)$ by computing

$$\hat{p}_n(X_i) = \Pr(D_i = 1|X_i = x)$$

2. Use the magic formula to estimate $E(TE(x)|x \in A)$: That is,

$$\frac{\sum_{D_i=1,x_i\in A} a_i Y_i}{\sum_{D_i=1,x_i\in A} a_i} - \frac{\sum_{D_i=0,x_i\in A} b_i Y_i}{\sum_{D_i=0,x_i\in A} b_i}$$

where $a_i = \frac{1}{\hat{p}_n(x_i)}, b_i = \frac{1}{1-\hat{p}_n(x_i)}$. The above can also be written as

$$\frac{\sum_{i=1,x_i\in A}^{N} \frac{D_i Y_i}{\hat{p}(X_i)}}{\sum_{i=1,x_i\in A}^{N} \frac{D_i}{\hat{p}(X_i)}} - \frac{\sum_{i=1,x_i\in A}^{N} \frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}}{\sum_{i=1,x_i\in A}^{N} \frac{1-D_i}{1-\hat{p}(X_i)}}$$

The reason this works is as follows: Notice that

$$\begin{aligned} D_i Y_i &= D_i(D_i Y_i(1) + (1-D_i)Y_i(0)) = D_i Y_i(1) \\ (1-D_i)Y_i &= (1-D_i)Y_i(0) \end{aligned}$$

Thus, we have

$$\begin{aligned} E\left[\frac{D_iY_i}{p(X_i)}\right] &= E\left[E\left[\frac{D_iY_i(1)}{p(X_i)}|X_i\right]\right] \\ &= E\left[\frac{E[D_i|X_i]E[Y_i(1)|X_i]}{p(X_i)}\right] \\ &= E\left[\frac{p(X_i)E[Y_i(1)|X_i]}{p(X_i)}\right] = E[E[Y_i(1)|X_i]] = E[Y_i(1)] \end{aligned}$$

so we have $E\left[\frac{D_iY_i}{p(X_i)}\right] = E[Y_i(1)], E\left[\frac{D_iY_i}{p(X_i)}|X_i\right] = E[Y_i(1)|X_i]$. We can make the similar arguments for $Y_i(0)$ and get $E\left[\frac{(1-D_i)Y_i}{1-p(X_i)}\right] = E[Y_i(0)], E\left[\frac{(1-D_i)Y_i}{1-p(X_i)}|X_i\right] = E[Y_i(0)|X_i]$. Thus, the ATE for $X_i \in A$ can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_iY_i}{p(X_i)} - \frac{(1-D_i)Y_i}{1-p(X_i)}\right) \ \forall X_i \in A$$

In most cases, the propensity score is estimated. So we use the estimator involving the $\hat{p}$, which is what we have in the lecture notes. It is said that normalizing the weights to one in finite samples improves the mean squared error properties of the estimator (Imbens, Rubin (2019) - *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*)

2.6 Matching Estimation

The idea behind matching estimation is to impute the values for something we cannot get from the data: Namely, $Y_i(0)$ for those in $D_i = 1$ and $Y_i(1)$ for those in $D_i = 0$. In other words, you try to find a 'close' match for a particular unit i in a different treatment arm. When we say we find k-closest neighbors for unit i in $D_i = 0$, we find k individuals in $D_i = 1$ that has close traits (X_i) to individual i, or k individuals with the smallest values of $||x_j - x_i||$. Then, we construct a counterfactual $Y_i(1)$ by taking a (weighted) average over the Y_i's of the k individuals found in the other group. The treatment effect would than be $Y_i(1)$ -Y_i, where $Y_i(1)$ is calculated as in previous sentence.

Let's try with $k = 1$ - we find the one individual in the opposite treatment arm that has the similar value of X_i. Then, the average treatment effect can be written as

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i(1) - \hat{Y}_i(0))$$

where

$$\hat{Y}_i(1) = \begin{cases} Y_i & (D_i = 1) \\ \text{imputed value} & (D_i = 0) \end{cases}, \ \hat{Y}_i(0) = \begin{cases} \text{imputed value} & (D_i = 1) \\ Y_i & (D_i = 0) \end{cases}$$

There are some caveat in this approach

- The covariate X_i that is used to find the closest match in the other treatment arm should not affect the assignment of the treatment. In other words, if the assignment of treatment vs. control depends on some subset of X_i, you should not use that subset as a standard for finding the closest match.
- The overlap condition becomes critical. If it is not satisfied, i.e. for some X_i, $p(X_i) = 1$ or 0, then we are unable to find a closest match in the other treatment arm because for individuals with that covariate value, all of them are either in control or treatment group and not spread around.
- Who are the neighbors? The answer might depend on what distance measure we use. Moreover, how many of those in the treatment can be considered neighbors? There is a trade-off in the sense that using larger k would force us to put someone who is not 'close' as neighbors and using small k may cause difficulty in imputing counterfactual values.

2.7 Difference-in-Difference Estimation

This involves a specific framework where we can clearly define a 'before and after' - denoted as t_0 and t_1. No one is treated at t_0 but there is a subset of people ($G_i = 1$) that are treated at t_1. Those in $G_i = 0$ are never treated in either time period. If we define

$$Y_{i,t} = G_i Y_i(1,t) + (1 - G_i) Y_i(0,t) \ \ (t \in \{t_0, t_1\})$$

and impose

$$Y_i(G_i, t_0) = Y_i(t_0) \ \text{ for both } G_i = 1 \text{ and } G_i = 0$$

then we will be able to observe $(Y_{i,t_1}, Y_{i,t_0}, G_i.X_i)$ for every unit i. What we do not observe is $Y_i(1, t_1)$ for those in $G_i = 0$ and $Y_i(0, t_1)$ for those in $G_i = 1$. The analogue to the conditional independence assumption in this context is a **parallel trend assumption**, defined as

$$(Y_i(1,t_1) - Y_i(t_0), Y_i(0,t_1) - Y_i(t_0)) \perp\!\!\!\perp G_i|X_i$$

To see why they are equal, consider a setting where $D_i = G_i$ and write

$$\begin{aligned} Y_i = Y_{i,t_1} - Y_{i,t_0} &= D_i(\underbrace{Y_i(1,t_0) - Y_i(1,t_0)}_{Y_i(1)}) + (1-D_i)(\underbrace{Y_i(0,t_1) - Y_i(0,t_0)}_{Y_i(0)}) \\ &= D_iY_i(1) + (1-D_i)Y_i(0) \end{aligned}$$

Therefore, we can rewrite the parallel trend assumption as

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i|X_i$$

which is the same as the conditional independence assumption.

To apply this in regression, we can write

$$Y_{it} = \beta_0(X_i) + \beta_1(X_i)\cdot 1(t=t_1) + \beta_2(X_i)\cdot G_i + \beta_3(X_i)\cdot G_i \cdot 1(t=t_1) + \epsilon_{it}$$

where X_i is a set of covariates, which can include a constant. In this context, the treatment effect for $X_i = x$ would be

$$\begin{aligned} TE(x) &= E[Y_i(1) - Y_i(0)|X_i = x] \\ &= E[(Y_i(1,t_1) - Y_i(1,t_0)) - (Y_i(0,t_1) - Y_i(0,t_0))|X_i = x] \\ &= x \cdot E\{[(\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2)] - [(\beta_0 + \beta_1) - (\beta_0)]|X_i = x\} \\ &= x \cdot E[(\beta_1 + \beta_3) - \beta_1|X_i = x] \\ &= \beta_3 x \end{aligned}$$

So β_3 would be our parameter of interest.

One difficulty arises from testing the parallel trends assumption. Loosely speaking, what we can do is to select a time period $\tilde{t} < t_0$ and find out if the difference $y_i(t_0) - y_i(\tilde{t})$ is independent with G_i. If not independent, we say that there is an autonomous trend.