Consider The Linear Probability Model Y = Bo+B1X; +ui, Where Pr(Y; = 1X) = Bo+B1Xi.(a) Show That E(u,

Consider The Linear Probability Model Y = Bo + B1X + ui, Where Pr(Yi = 1|Xi) = Bo + B1Xi. (a) Show That E(ui) = 0

Introduction

The Linear Probability Model (LPM) is a foundational tool in econometrics and statistics for modeling binary dependent variables. It expresses the probability that an event occurs as a linear function of independent variables. Although simple and intuitive, the LPM has certain limitations, such as predictions outside the [0,1] interval and heteroskedasticity issues. Nevertheless, it provides valuable insights into the relationship between explanatory variables and binary outcomes.

In this article, we focus on a specific aspect of the Linear Probability Model: demonstrating that the expected value of the error term (ui) is zero, i.e., E(ui) = 0. This property is crucial for understanding the assumptions underlying the Ordinary Least Squares (OLS) estimator and ensuring the unbiasedness of estimates within the model.

We begin with a brief overview of the model's structure, followed by a step-by-step demonstration of why the expectation of the error term is zero under the model assumptions. This discussion is not only theoretically important but also provides practical insights for researchers applying the LPM in empirical analyses.

Understanding the Linear Probability Model (LPM)

Model Specification

The LPM is specified as:

$Y_i = Bo + B1X_i + u_i$

where:

- **Y_i**: Binary dependent variable, taking values 0 or 1.
- **X_i**: Independent variable(s), which can be continuous or discrete.
- **Bo**: Intercept term.
- **B1**: Slope coefficient(s) indicating the effect of X on Y.
- **u_i**: Error term capturing unobserved factors and model deviations.

Probability Interpretation

The key assumption of the LPM is that the probability that Y_i equals 1, given X_i, is a linear function of X_i:

$Pr(Y_i = 1 \mid X_i) = Bo + B1X_i$

Since probabilities must lie within [0,1], the linear function is often restricted or used with caution, especially when predicted probabilities fall outside this range.

Deriving the Expectation of the Error Term E(ui)

Step 1: Expressing the Error Term

By the model specification, the actual observed value Y_i can be written as:

$Y_i = Bo + B1X_i + u_i$

Rearranged, the error term is:

$u_i = Y_i - (Bo + B1X_i)$

Step 2: Computing the Conditional Expectation $E(u_i \mid X_i)$

To analyze E(ui), we consider the conditional expectation given X_i:

$E[u_i \mid X_i] = E[Y_i - (Bo + B1X_i) \mid X_i] = E[Y_i \mid X_i] - (Bo + B1X_i)$

From the model assumption, the probability that Y_i equals 1, given X_i, is:

$Pr(Y_i = 1 \mid X_i) = Bo + B1X_i$

Since Y_i is a Bernoulli random variable:

$E[Y_i \mid X_i] = Pr(Y_i = 1 \mid X_i) = Bo + B1X_i$

Step 3: Demonstrating that $E(u_i \mid X_i) = 0$

Substituting back into the expression for the conditional expectation:

$E[u_i \mid X_i] = (Bo + B1X_i) - (Bo + B1X_i) = 0$

This shows that, conditioned on X_i, the expectation of the error term is zero.

Unconditional Expectation E(ui) = 0

Step 4: Taking the Law of Total Expectation

The unconditional expectation of ui is obtained by applying the law of total expectation:

$E[u_i] = E[\ E[u_i \mid X_i]\ ]$

From the previous step, we know that $E[u_i \mid X_i] = 0$ for all X_i. Therefore:

$E[u_i] = E[0] = 0$

Implication of the Result

This result indicates that, under the model's assumptions, the error term has a zero mean over the entire sample. This zero-mean property is essential because it ensures the unbiasedness of the OLS estimators for the coefficients Bo and B1, provided other standard assumptions (like independence, homoscedasticity, etc.) hold.

Assumptions Underpinning E(ui) = 0

Key Assumptions for the Result

1. **Correct Model Specification:** The model correctly specifies the relationship between the probability of Y=1 and X.

2. **Linearity of the Conditional Probability:** The probability $Pr(Y_i = 1 \mid X_i)$ is linear in X_i.

3. **Independence of Error Term and X_i:** The error term u_i has an expected value of zero conditional on X_i.

4. **Exogeneity of X_i:** The explanatory variables are uncorrelated with the error term.

Violations of these assumptions, especially the independence assumption, can lead to biased or inconsistent estimates, which is why the LPM has limitations in practical applications.

Conclusion

In summary, the demonstration that E(ui) = 0 in the Linear Probability Model is rooted in the model's assumption that the expected value of the binary dependent variable Y_i, conditioned on the independent variable X_i, equals the linear function Bo + $B1X_i$. By expressing the error term as the difference between the observed Y_i and its conditional expectation, and then applying the law of total expectation, we establish that the mean of the error term across the entire sample is zero.

This property is fundamental in ensuring the unbiasedness of OLS estimates in the LPM, although practitioners should remain cautious of the model's limitations, especially regarding predicted probabilities outside the [0,1] range and heteroskedasticity.

Understanding the expectation of the error term provides a solid foundation for more advanced binary choice models, such as the Logistic and Probit models, which address some of the limitations inherent in the Linear Probability Model.

Frequently Asked Questions

What is the Linear Probability Model (LPM) in the context of binary outcomes?

The Linear Probability Model (LPM) is a regression model where the dependent variable is binary (0 or 1), expressed as $Y = \beta_0 + \beta_1 X + u$, aiming to estimate the probability that Y equals 1 given X.

In the LPM, how is the probability Pr(Y=1|X) expressed?

$Pr(Y=1|X) = \beta_0 + \beta_1 X$, which is a linear function of the independent variable

X.

What does the error term u_i represent in the LPM?

The error term u_i captures the deviation of the observed outcome from the linear probability predicted by $\beta_0 + \beta_1 X$, accounting for unobserved factors and randomness.

How do we demonstrate that E(u_i | X) = 0 in the LPM?

Since the expected value of Y_i given X is the probability Pr(Y=1|X), and the model predicts this probability as $\beta_0 + \beta_1 X$, the residual $u_i = Y_i - (\beta_0 + \beta_1 X)$. Taking expectations conditional on X, $E(u_i | X) = E(Y_i | X) - (\beta_0 + \beta_1 X) = Pr(Y=1|X) - Pr(Y=1|X) = 0$.

Why is the assumption E(u_i | X) = 0 important in the linear probability model?

This assumption ensures that the model is correctly specified and that the error term is uncorrelated with the independent variables, allowing for unbiased and consistent estimates of the coefficients.

What are some limitations of the Linear Probability Model?

The LPM can predict probabilities outside the [0,1] range and assumes constant variance of errors (heteroskedasticity), which can lead to inefficiency and inaccuracies in estimation.

How does the LPM relate to the concept of expected value of the error term?

In the LPM, the expected value of the error term conditioned on X is zero, $E(u_i | X) = 0$, which indicates that on average, the residuals are centered around zero for given X.

Can the LPM be used for causal inference?

While the LPM can provide insights into relationships between variables, its limitations mean that for causal inference, more appropriate models like logistic regression are often preferred due to better handling of the binary nature of the outcome.

What is the main takeaway regarding the expectation

of the error term in the linear probability model?

The main takeaway is that, under proper model specification, the expected value of the error term conditioned on X is zero, i.e., $E(u_i | X) = 0$, which is fundamental for the unbiasedness of the Ordinary Least Squares estimators.

Additional Resources

Consider The Linear Probability Model Y = Bo + B1X + u_i is a fundamental concept in econometrics and statistical modeling, especially when dealing with binary dependent variables. The linear probability model (LPM) offers a straightforward approach to estimate the probability that an event occurs, given a set of explanatory variables. Despite its simplicity and intuitive appeal, understanding the properties and limitations of the LPM is crucial for analysts, researchers, and students alike. In this comprehensive guide, we will delve deeply into the derivation, interpretation, and implications of the model, focusing specifically on demonstrating that $E(u_i | X) = 0$, a key assumption underpinning the validity of the model.

Introduction to the Linear Probability Model

The Linear Probability Model (LPM) is a type of linear regression where the dependent variable, Y, is binary — taking values of 0 or 1. The model is specified as:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where:

- Y_i indicates the occurrence (1) or non-occurrence (0) of an event,
- X_i is a vector of independent variables (or a single independent variable),
- β_0 and β_1 are parameters to be estimated,
- u_i is the error term capturing the deviation of observed Y_i from its expected value.

The model estimates the probability that $Y_i = 1$ as a linear function of X_i:

$$P(Y_i=1|X_i) = \beta_0 + \beta_1 X_i$$

The Core of the Model: Showing $E(u_i|X_i) = 0$

One of the foundational assumptions in classical linear regression, including the LPM, is that the error term has an expected value of zero conditional on the explanatory variables:

$$ E(u_i|X_i) = 0 $$

This assumption ensures that the estimated coefficients are unbiased and consistent. To understand why this holds in the context of the LPM, let's explore the derivation step by step.

Step 1: Recognizing the Nature of Y_i

Since Y_i is binary:

$$ Y_i =
\begin{cases}
1, & \text{if the event occurs} \\
0, & \text{if the event does not occur}
\end{cases}
$$

The expected value of Y_i given X_i is:

$$ E[Y_i|X_i] = P(Y_i=1|X_i) = \beta_0 + \beta_1 X_i $$

This is the conditional mean of Y_i.

Step 2: Expressing the Error Term

Rearranging the model:

$$ u_i = Y_i - (\beta_0 + \beta_1 X_i) $$

The error term captures the difference between the actual observed value and the model's predicted probability.

Step 3: Calculating the Conditional Expectation of u_i

Taking the expectation of u_i conditional on X_i:

$$
\begin{aligned}
E[u_i|X_i] &= E[Y_i - (\beta_0 + \beta_1 X_i) | X_i] \\
&= E[Y_i|X_i] - (\beta_0 + \beta_1 X_i) \\
&= P(Y_i=1|X_i) - (\beta_0 + \beta_1 X_i)
\end{aligned}
$$

Since the model specifies:

$$ P(Y_i=1|X_i) = \beta_0 + \beta_1 X_i $$

it follows that:

$$ E[u_i|X_i] = (\beta_0 + \beta_1 X_i) - (\beta_0 + \beta_1 X_i) = 0 $$

This derivation confirms that the error term has a zero mean conditional on X_i, satisfying a key assumption for unbiased Ordinary Least Squares (OLS) estimates.

Implications of the Zero Conditional Mean Assumption

The conclusion that $E[u_i|X_i] = 0$ in the linear probability model has several important implications:

1. Unbiased Estimators

When the assumption holds, the OLS estimators for β_0 and β_1 are unbiased, meaning on average, they hit the true parameter values across many samples.

2. Consistency of Estimators

The estimators are also consistent, converging in probability to the true parameters as the sample size increases.

3. Interpretability

Because the model is linear in parameters, the coefficients are directly interpretable as the change in probability associated with a one-unit change in X_i.

Limitations and Challenges of the Linear Probability Model

While the assumption $E[u_i|X_i] = 0$ is crucial, the LPM faces several practical issues:

1. Predicted Probabilities Outside [0,1]

Since the model is linear, it can produce predicted probabilities less than 0 or greater than 1, which are not meaningful probabilities.

2. Heteroskedasticity

The variance of u_i depends on X_i, violating the homoskedasticity assumption:

$$\operatorname{Var}(u_i|X_i) = P(Y_i=1|X_i)(1 - P(Y_i=1|X_i))$$

which varies with X_i, leading to inefficient estimates.

3. Violation of the Error Term Assumption

The key assumption $E[u_i|X_i] = 0$ hinges on the correct functional form and the absence of omitted variables that are correlated with X_i. If these are present, bias can occur.

Alternatives and Extensions

Given these limitations, econometricians often prefer other models for binary dependent variables:

- Logit Model: Uses a logistic function to constrain predicted probabilities within [0,1], ensuring meaningful probabilities.
- Probit Model: Employs a cumulative normal distribution function for probability estimation.
- Linear Probability Model with Robust Standard Errors: Sometimes used as a simple approach with adjustments for heteroskedasticity.

Summary of Key Points

- The Linear Probability Model expresses the probability of an event as a linear function of predictors.
- The error term u_i captures deviations from the predicted probability.
- The fundamental assumption $E[u_i|X_i] = 0$ holds because the model explicitly specifies the conditional mean of Y_i.
- This assumption underpins the unbiasedness and consistency of OLS estimates in the LPM.
- Despite its simplicity, the LPM suffers from limitations like predicted probabilities outside [0,1], heteroskedasticity, and potential bias if model assumptions are violated.

Final Thoughts

Understanding the derivation that $E(u_i|X_i) = 0$ in the linear probability model is essential for appreciating its strengths and limitations. It emphasizes that, under correct specification, the model provides a straightforward way to estimate how variables influence the likelihood of an event. However, econometricians must be cautious and consider alternative models when the assumptions are questionable or the model produces unrealistic predictions. Mastery of these concepts aids in making informed decisions about which modeling approach best suits their data and research questions.

Remember: The linear probability model offers an intuitive starting point,

```
but always consider its assumptions carefully and explore more advanced
models for binary outcome analysis when appropriate.
```

Consider The Linear Probability Model Y Bo B1x Ui Where Pr Y 1x Bo B1xi A Show That E U

Consider The Linear Probability Model Y Bo B1x Ui Where Pr Y 1x Bo B1xi A Show That E U

Related Articles

- Find All The Female Characters From The Series "aot" Who Have "ann" In Their Name In One Single Exact
- Based On Your Knowledge Of The Geography Of Ancient Mesopotamia Choose The Two Cities On The Map That
- Determine The Type Of Inhibition That Has Occurred In A First Order Michaelis-Menten Enzyme Catalyzed

Back to Home