Adjusting for Endogeneity in the AFT Model

Muhammad Atiyat (Advisor: Dr. Debashis Ghosh)
Department of Statistics, Penn State University

1. Event Time Data
- Interest is in the time until a particular event occurs.
- Often this data is subject to censoring.
- Censoring refers to the fact that we are unable to observe all event times by the end of our study.
- Could be due to a subject: not experiencing the event of interest by the end of the study, experiencing another event, or dropping out of the study.

2. The Accelerated Failure Time (AFT) Model
- The AFT model is one of the well known models used to examine relationships between an event time subject to censoring and explanatory variables.
- It expresses a linear relationship between the event time of interest and the explanatory variables:
\[ T_i = X_i \beta + \epsilon_i \] (1)
where \( T_i \) is a scalar response variable, which represents the logarithm of the event time of interest for subject \( i \), \( X_i \) is a \( 1 \times p \) vector of explanatory variables, \( \beta \) is a \( p \times 1 \) vector of unknown regression coefficients, \( \epsilon_i \) is a random variable with an unspecified distribution with zero mean and a finite variance.
- Due to censoring we do not observe \( T_i \). Instead we observe \( Y_i = \min(T_i, C_i) \), where \( C_i \) is a censoring time, and \( \delta_i = I(T_i \leq C_i) \) is an indicator of the event.
- Two known methods used to estimate the AFT model are by Buckley and James (1979) and Koul et al. (1981). These two methods are referred to by BJ and KSV respectively.

3. Endogeneity
- The diagram above illustrates that the explanatory variables \( X_i \) are endogenous because they depend on unobserved variables which in turn affect the event time of interest.
- The use of instruments is one way to adjust for endogeneity. An instrument must: be correlated with the endogenous explanatory variables, not be correlated with the unobserved confounders, and not have a direct influence on the event time.
- Angrist et al. (1999) come up with two methods, J1 and J2, which use instruments to adjust for endogeneity in a linear model which is similar to the AFT model but does not accommodate for censoring.

4. Proposed Methods
- Goal: adjust for endogeneity in the AFT model.
- To accomplish this, we propose combining the methodology of Buckley and James (1979) and Koul et al. (1981) with that of Angrist et al. (1999).
- We propose three methods: J1KSV, J1MKSV, and J1BJ. Note: MKSV is similar to KSV but with a slight modification.

5. Simulation Study Results
- Using simulation studies, we compare the estimation of 8 methods: J1, J2, KSV, J1KSV, MKSV, J1MKSV, BJ, and J1BJ where J1 and J2 adjust for endogeneity only, KSV, MKSV, and BJ adjust for censoring only, and J1KSV, J1MKSV, and J1BJ adjust for both endogeneity and censoring in Model (1).
- 1000 simulation samples of \( n=1000 \) each and \( \approx 51\% \) censoring.
- Here the AFT model consists of 3 parameters: the intercept \( \alpha \), the coefficient \( \beta_1 \) of an endogenous categorical variable, and the coefficient \( \beta_2 \) of an endogenous continuous variable.

6. PBC Data
- Primary biliary cirrhosis (PBC) is a chronic liver disease associated with the irritation and swelling of the liver’s bile ducts.
- Event of interest: death. A total of 312 patients and 125 deaths. The analysis is restricted to a total of 163 patients and is \( \approx 50\% \) censoring.
- Variable bilirubin is the level of serum bilirubin (mg/dl), a yellow pigment found in bile, in one’s blood. It is measured at followup times, after treatment assignment, and continues to be measured until the patient is censored or dead. Therefore, it may provide information about one’s death.
- bilirubin is endogenous because it is related to unobserved indicators of a patient’s health which in turn affect death.

7. Data Analysis
- Instrument: bilirubin taken at 1 year minus bilirubin at baseline.
- Explanatory variables: bilirubin, age, and trt, i.e., treatment.
- Results are shown below. Values in ( ) represent standard errors based on a simple bootstrap sample of 500.

8. Results and Conclusions
- The estimates of bilirubin become negative when adjusting for endogeneity for KSV and MKSV.
- The estimates of the coefficients of age and trt are quite different depending on the method.
- This is an indication that adjusting for endogeneity in the presence of censoring may be important in the PBC data depending on the method which is used.

References