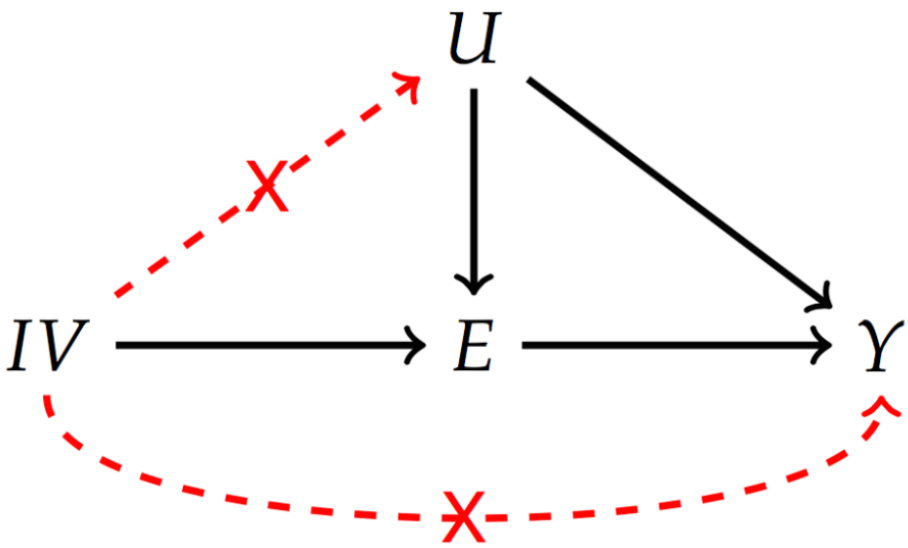
Discuss the exogeneity assumption in regression analysis. Provide examples for its possible violation and strategies to address them.

As per Glen (2018), exogeneity shows the lack of dependence of X (Independent Variable) on Y (Dependent Variable). The connection may be present but would still depend on IV and error term. Exogneity in relation to a variable implies that variable is not affected by the presence of other variables, thus exogeneity implies that the variable would influence other variables but not get influenced in turn. The strictly exogenous implies that error term is unrelated to variable X and completely unaffected by Y. On the other hand, sequentially exogenous variable or predetermined variable means that error term is unrelated to variable X instances – if variable does not fall in either of these conditions, then it is called endogenous equation or variable. An example would be building a model to describe effect of weather on playing football in a town in the east. Weather is independent variable and is affecting the probability of the sport play. The relationship does not work the other way and sport play probability would not affect the weather – hence weather is an exogenous variable (Glen, 2018).

Ben (2008) states that OLS (ordinary least squares) assume that the errors in regression are uncorrelated with the dependent variable. He suggests that TSLS (Two Stages Least Estimators) can provide a number of useful tests for Exogeneity/excludability of instruments checking if a correlation exists between DV and variation in variable exists and if that variation can be explained exogeneity. As per Han (2011), OLS assumption of exogeneity is only applicable when $E(u|X) = 0$ and no correlation exists between X and u.

Professor (Fingleton, n.d.) explains that for exogeneity X variable is not dependent on the DV and Y depends on X and error e, this X is assumed to be independent of Y, if this condition is not met, the estimated coefficients for regression would not be consistent when OLS estimating is

done. If the OLS estimation is unbiased, the explanatory variables should not be correlated to the error term. Exogeneity states that instrument is not correlated with the error term and y should only be affected through x as shown here



(Baumann, 2020)

The *log(wages)* = $\alpha + \beta Edu + \varepsilon$ equation may show correlation between Edu and error term because of family background and ability of the person. To ensure relevance and exogeneity, the independent variable should be uncorrelated with the error term and predict Edu. TSLS is used for finding a new Edu that could then be used in the original equation

$$1) \quad \widehat{Edu}_i = \hat{\delta}_0 + \hat{\delta}_1 Fam_i + V_i$$

$$2) \quad \log(\widehat{wage}_i) = \hat{\beta}_0 + \hat{\beta}_1 \widehat{Edu}_i + \varepsilon_i$$

(Baumann, 2020)

Instrumental variable of family background is computed and then substituted in the initial equation to update the initial wage estimates. This estimate using IV is expected to be unbiased and reduced the bias issue (Baumann, 2020).

Estimating a model using OLS requires inclusion of relevant variables without which the causal effect may appear to be exogenous but the results may be erroneously associated with the present variables. Thus, the estimated effect of marketing variable on the DV would be biased and distorted because independent variables and unobserved factors, included through error term are correlated.

Homburg, Klarmann and Vomberg (2017) explain that regression analysis estimates causal properties of regressor on the DV with the assumption of correlation between regressor and error term. Endogeneity is addressed by finding additional variables called instrumental variables that are correlated with regressor but not the error term. Instrumental variables must be strongly related to the endogenous regressors and should be unrelated to the error term. As explained by the authors, marketing is concerned with the estimation of regression models for understanding effects of prices or investments on gaining market shares or increasing sales. Causal relationships determination may be hampered by endogeneity in the observational data. The independent variables may be strategically set but may not be observable, resulting in the unobserved factors being part of error term making the independent variables biased for OLS estimation. Handling such issues requires using exogenous instrumentation variables that capture the unobserved factors. Field experiments can help in estimating exogenous variations. In the field experiments the independent variables values are randomly set and on the basis of resultant effects on the dependent variable, the causal effects can be gauged via OLS where randomization of independent variable ensures its exogeneity (Homburg, Klarmann and Vomberg, 2017).

As stated by Abdallah, Goergen and O'Sullivan (2015) endogeneity is a major issue across management, and business fields, strategic decisions and, consequently, organizational performance can be affected. First step in addressing endogeneity is determining its source, if it is because of omitted variables in the equation then these could be included.

$$E(u_t|x_{t1}, \ldots, x_{tk}) = E(u_t|\mathbf{x}_t) = 0$$

When the above equation holds that x_{ij} are contemporaneously exogenous. Equation implies that u_i and the explanatory variables are uncorrelated with x_{sj} even when $s \neq t$ the explanatory variables must be exogenous. OLS can be unbiased only when exogeneity assumption holds. As Wooldridge (2009) explains, explanatory variables which are clear exogenous would not react to y values of the past. An example could be how much rainfall was received in the agricultural field – subsequent year's production is not influenced by agricultural output in the prior years. Labor used might however not be exogenous as the labor input would be dependent on previous year's crop yields. Similarly the policy variables like money supply and speed limits on freeways, welfare expenditures are all linked to previous years' outcomes.

Strict application of exogeneity would mean that the error at time t, u_t, is not correlated with the regressed variables in all the time periods evaluated: in the current time, earlier and in the future. When sequential exogeneity is implied, it means that u_t is uncorrelated with current and earlier used variables as regressors, so is shown by strict exogeneity. Also, qequential exogeneity means that u_t is uncorrelated with x_t, x_{t-1}, …, implying that u_t is not correlated with x_t (Wooldridge, 2009).

Select one article using the event study methodology. Discuss how the event study methodology is implemented and used to test the hypotheses or answer the research questions of the article.

The article used for this section is Stock liquidity and default risk written by Jonathan Brogaard, Dan Li , Ying Xia and published in Journal of Financial Economics (2017). This paper addresses the corporate default risk reduction through liquidity in stocks. Authors believe that stock liquidity is related to the risk of default as it may lead increased monitoring or it may lead to greater mispricing; but on the positive side, having more liquidity might decrease the risk of default as that would improve the price efficiency and ensure better corporate governance. The net impact of liquidity in stocks depends on several factors. It is therefore important to understand the specific relationship between the liquidity and risk of default.

Bharath and Shumway (2008) model of measuring expected default frequency (EDF) is used as this model has proven to be accurate in predicting bankruptcies. US equity market between 1994 and 2014 is used for the analysis. Examining liquidity requires measuring bilateral relationship between liquidity and default risk. Causality (reverse) is addressed by difference-indifferences (DID) analysis for time era of 2001 where SEC had implemented tick size change from sixteenth of dollar to hundredth of dollar (decimalization). Earlier studies had shown improvement in liquidity when decimalization was implemented. EDF decreased by 6.2% around decimalization with high risk stock experiencing the most decrease in EDF. Two possible avenues are considered that might drive the relation between having liquidity compared to the default risk, which are the efficiency of information and the governance. Better quality of information would allow investment managers to make more informed decisions and good governance ensure that

managers engage in investments that enhance value and avoid riskier investments, thus reducing default probability.

The sample constitutes US stocks appearing in both Compustat and Center for Research in Security Prices (CSRP) after excluding firms that were subject to statutory capital requirements. The sample included organizations with more than 200 active trading days per year. Financial data is obtained from Compustat and CSRP databases. Both low and high frequency measures are used to capture stock liquidity. Liquidity is guaged by relative effective spread measure expressed as a percentage value. Daily effective spread is calculated by taking weighted average of the effective spread records. Low liquidity is implied by high spread. Liquidity is also measured by percentage quoted spread. For the low frequency measures, Amihud illiquidity ratio captures stocks that experienced larger change in prices and hence are more illiquid. Measures, *zeros* is defined as the proportion of total days where stocks had zero returns as the illiquid stocks would have more zero returns, thus higher value shows lower liquidity. All of these four measures were negatively correlated with liquidity and default risk. Multivariate analysis shows bilateral relation between liquidity and default risk. Market makers typically provided lesser liquidity for stocks with higher default risks. Tick price decimalization is incorporated as an exogenous shock to the liquidity in stocks. Decimalization allowed having tighter spreads between the bid and the ask prices and thus helped to improve liquidity amongst actively traded stocks. Focusing on year before the event and year after the event decimalization took place helped to spot out unobserved variables. Difference in Difference (DID) strategy is used to compare the change in default probability for two firm groups that were similar, except for experiencing significant change in liquidity when decimalization took place. DID controlled for omitted and unobserved variables. Constructing a treatment and a control group through

matching propensity scores and ranking all sampled firms three groups of firms were constructed and the first and the third group of companies were retained. Through using a probit model that set a value of 1 for all of the firms in the group one and 0 for the firms in the group three and by including the effective spread, the firms in the two groups were matched. Probit model included effective spread and included the controls as specified earlier for both pre decimalization and post decimalization periods. Propensity scores with 0.01 are used to match firms between the first and third groups. The treatment and control pairs model had a pseudo R^2 value of 0.1068 and $p<0.001$ thus showing that model captured significant variation of the choice variable. Stock liquidity effect would be highest as firm is about to default, so matched pairs are grouped to test the hypothesis. Increasing price efficiency and better corporate governance were both found to be the driver for the results. Using both measures in regression equation standardized regression equation was developed. The DV and independent variables were subtracted by average values and resulting difference divided by STD[1].

Coefficient can be read as the impact that one STD change in independent variable on the DV. The price efficiency channel was shown to have better explanatory magnitude compared to the corporate governance channel. Change of one STD in delta for correlation (change in price delay) was found to lead to EDF change of 11% to 12 %. The ability/magnitude of explaining the governance channel was found lower than price efficiency channel.

Negative relation existed between stock liquidity and bankruptcy risks. Decimalization event was used as an exogenous shock to use DID for examining negative effect of stock liquidity on default propensity of firms. The impact was found to be higher in pre-decimalization period. Increasing stock liquidity was shown to reduce default risk, and decreasing the minimum tick

[1] STD = Standard Deviation

size for stock can decrease default risks for organizations. This article examines the decimalization as an event that signifies changes in default risks but other market designs need to be examined to check effect on default risk and possible comparison with other markets would help in isolating most important factors and possibly identify other factors that impact default propensity that were not evident in US equity market.

References

Abdallah, W., Goergen, M. and O'Sullivan, N., 2015. Endogeneity: How Failure to Correct for it can Cause Wrong Inferences and Some Remedies. *British Journal of Management*, 26(4), pp.791-804.

Baumann, D., 2020. *Instrumental Variables: A practical explanation*. [online] Medium. Available at: <https://towardsdatascience.com/instrumental-variables-a-practical-explanation-1a583408a5b9> [Accessed 2 May 2021].

Brogaard, J., Li, D. and Xia, Y., 2017. Stock liquidity and default risk. *Journal of Financial Economics*, 124(3), pp.486-502.

Fingleton, B., n.d. *Endogeneity, Exogeneity and instrumental variables*. [online] Cantab.net. Available at: <http://www.cantab.net/users/bf100/pdf/iv_slides_fingleton.pdf> [Accessed 2 May 2021].

Glen, S., 2018. *Exogeneity: Definition - Statistics How To*. [online] Statistics How To. Available at: <https://www.statisticshowto.com/exogeneity/> [Accessed 1 May 2021].

Han, E., 2011. *Econometrics Honor's Exam Review Session*. [online] Scholar.harvard.edu. Available at: <https://scholar.harvard.edu/files/economics/files/final_review_1123.pdf> [Accessed 2 May 2021].

Homburg, C., Klarmann, M. and Vomberg, A., 2017. *Handbook of Market Research*. Springer International Publishing.

Shepherd, B., 2008. *ARTNeT Capacity Building Workshop for Trade Research: "Behind the Border" Gravity Modeling*. [online] Artnet.unescap.org. Available at: <https://artnet.unescap.org/tid/artnet/mtg/gravity09_tues3.pdf> [Accessed 2 May 2021].

Wooldridge, J., 2009. *Introductory econometrics.* 4th ed. Natorp Boulevard, Mason OH: South-Western Cengage Learning.