JURNAL ILMIAH TEKNIK INDUSTRI
ISSN: 1412-6869 (Print), ISSN: 2460-4038 (Online)
Journal homepage: http://journals.ums.ac.id/index.php/jiti/index
doi: 10.23917/jiti.v22i2.23020

Causal Inference to Predict Delayed Arrival of Ordered Production Materials at PT. XYZ

Muhammad Ridzky Hanura[1a], Yusuf Priyandari[1b◆], Muhammad Hisjam[1c]

Abstract. *PT XYZ has a problem with the delayed arrival of ordered production materials. Although the company is aware of the delays based on data, the company does not yet know the causes or sources of problems that cause delays. On the other hand, not all factors can be controlled to reduce the delay in the arrival of production materials. The company intends to predict the change in delay time if control or intervention is carried out on certain factors by utilising data availability. The factor to be treated is requisition-to-order lead time A causal inference model is used using the Dowhy library (a Python library for causal inference by graphing the model, quantitatively evaluating causal effects, and validating the causal assumptions) to estimate the quantitative causal effect between requisition-to-order lead time and the arrival time of the ordered material by considering other factors that also affect the delay. The results of the causal effect estimation are that by intervening or controlling the requisition-to-order lead time factor by one day, there is a decrease in the average delay in material arrival time by one day.*

Keywords: *causal effects, causal inference, supply chain management*

I. INTRODUCTION

Supply chain management is managing all activities in the supply chain, starting from raw materials to semi-finished or finished products, then delivering these products to consumers through a distribution system (Heizer et al., 2020). Supply chain management is an important thing that companies must consider. One of the main activities in the supply chain is material procurement. Material procurement is one of the main components of supply chain management, where the procurement process can create advantages in terms of time (Pujawan & Mahendrawati, 2017).

PT XYZ is a company engaged in the manufacturing industry of hospital equipment products. PT XYZ produces various products, including patient beds, operating tables, and wheelchairs. The production material procurement process at PT XYZ starts with a Purchase Requisition (PR), which is made before making a Purchase Order; after the Purchase Requisition is made, the company makes a Purchase Order to the supplier, and there is an agreement on the date when the material must have arrived (delivery date). PT XYZ waits until the ordered material is received (good receipt). There is a lead time in the purchase requisition to purchase order and purchase order to good receipt. Lead time is an important factor most easily observed in assessing performance and every process (Putra & Vikaliana, 2022). In the material procurement process, the existing lead time always fluctuates. Fluctuating lead times cause problems in material procurement because fluctuating lead times can cause a decrease in supply chain performance in material procurement and also increase costs (Dolgui et al., 2013). Uncertainty in lead time is one of the main sources of risk in the supply chain (Colicchia et al., 2010). One of the supply chain risks in material procurement is the occurrence of delays. Delays occur when the material's arrival time exceeds what the company expects. Delays or material unavailability can cause losses to the company, for example, the cessation of the production process and products not being delivered to consumers, product delivery to consumers becoming late, and losses in worker

[1] Industrial Engineering Department, Faculty of Engineering, Sebelas Maret State University, Jl. Ir. Sutami, Kentingan, Surakarta, 57126, Indonesia

[a] email: ridzky.hanura007@gmail.com
[b] email: priyandari@staff.uns.ac.id
[c] email: hisjam@staff.uns.ac.id
◆ corresponding author

Submited: 28-08-2023 Revised: 02-12-2023
Accepted: 24-12-2023

payments and other production costs (Kaban & Wicaksono, 2020).

PT XYZ has a problem with the delayed arrival of ordered production materials. This is supported by data stored in the Enterprise Resource Planning (ERP) system in 2022, showing a delay where the good receipt time exceeds the delivery date. Although the company knows there is a delay based on the data, the company does not yet know the causes or sources of problems that cause delays. Delays may be caused by the lead time between the PR and the PO, the quantity of material ordered, the ordering period in months, the supplier supply, the distance between the supplier and the company, and the type of supplier. On the other hand, not all factors can be controlled to reduce delays in the arrival of production materials. The company intends to predict the magnitude of the change in delay time if control or intervention is carried out on certain factors by utilising data availability.

Several methods can be used to predict material procurement delays using available data. Suppose the delay in material procurement is predicted based on numerical values. In that case, the method that can be used is the regression method. Although regression analysis can be used to predict the response variable using predictor variables, this method needs to improve where there is a risk of Simpson's Paradox. Simpson's Paradox is a statistical phenomenon where the relationship between two variables in a population appears, disappears, or reverses when the population is divided into several subpopulations. For example, two variables may be positively related in a population but independent or even negatively related in all subpopulations (Sprenger & Weinberger, 2021). At the same time, confounding factors also influence the two variables used in the regression analysis.

Based on these problems, a method called causal inference is used. Causal inference is a method to uncover the effects. It causes that influence something based on multivariate observational data, intervening ('What-if...") and counterfactual reasoning ("What if I do...") possible—the process of concluding based on conditions that existed at the time of the effect. Causality is generally considered the same as correlation. Still, there is a difference, where causality presents a different interpretation of observational data in the form of analysing asymmetric changes and responses between cause-and-effect variables (Chen et al., 2022).

Sun, et al. (2023) used causal inference for coastal water resources management in their research. Causal inference builds causal models that explain and evaluate groundwater problems by considering cause-and-effect relationships. Causal models have been successfully used to assess groundwater toxicity, hydrological patterns, and dependencies between variables. The causal inference was used to develop unbiased estimates of causal impacts on coastal aquifers, specifically the influence of groundwater pumping and river surface on chloride concentrations. The model built is a structural causal model using a Python programming language library called DoWhy (Sharma & Kiciman, 2020) and EconML (Oprescu, et al., 2019). The methods used for estimation are propensity score and double machine learning. After estimating, a refutation test is carried out to test the robustness of the estimation method. The method used for the refutation test is a placebo treatment.

Feng, et al. (2023), in their research, used causal inference to investigate the true causal relationship between air pollution, meteorological variables, and daily pediatric asthma patient visits (DPAPV) in Hangzhou, China, over the period 2014 to 2021. The study sought to determine if certain environmental variables caused an increase or exacerbation of pediatric asthma and if any short-term effects were involved. This study used the same model as the study by Sun, et al. (2023) with differences in the estimation method, which uses the Poisson method, and the refutation test method, which uses the method of adding a random common cause, placebo treatment, and uses a subset of data.

This paper aims to use causal inference to scope supply chain management problems by predicting changes in production material arrival delays with linear regression estimation methods and refutation test methods, such as adding a random common cause and using a subset of data.

II. Research Method

Data Collecting

The data collected is material procurement data for three hospital bed products. The data features collected are the material ordered, the material ordering period (in months), the order quantity, the supplier, the distance from the supplier to the company, the requisition-to-order lead time (the time between the date the purchase requisition is made and the date the purchase order is made), the supplier lead time (the time between the date the purchase order process is carried out and the date of agreement between the supplier and the company when the material has arrived at the company), and the good receipts lead time. Supplier lead time and good receipts lead time are used to measure the delay in ordered material arrival.

Causal Inference Analysis

The Dowhy library is a Python library for causal inference that graphs the model, quantitatively evaluates causal effects, and validates the causal assumptions. Four key steps were implemented: modelling, identification, estimation, and refutation.

A causal or directed acyclic graph is created in the modelling step. A directed acyclic graph is a visualisation form of a structural causal model that uses a directed graph and has no cycles. The DAG created then must be fully justified based on domain knowledge, theory, and research (Arif & MacNeil, 2023). After that, the variables that will become treatment and outcome variables are determined. By determining the treatment and outcome variables, confounder variables can be identified, which are variables that affect both treatment and outcome variables, and effect modifiers, which only affect the outcome variable.

In the identification and estimation step, the effect of the treatment is identified with the "backdoor criterion" (Pearl, 2009). This criterion is used to cut the backdoor path in the causal inference model, where the backdoor path is the path that affects the treatment and outcome variables. The effects of the identified treatments were estimated using the linear regression method after adjusting for confounder and modifiers effects. A significance test was conducted where p-value <0.05 indicates significance between the estimated equation and the outcome variable.

The estimation equation is tested for robustness to validate the estimation results in the refutation step. The methods used are adding a random common cause and using a subset of data. The adding a random common cause method is a method in the refutation test by adding synthetic independent random variables to the original data to verify the model's sensitivity to unobserved confounder variables. In the "subset of data" method, the original data are replaced by randomly selected subsets to evaluate the variance of the effect generated in the estimation step. The refutation test can also return a p-value where a p-value > 0.05 indicates that the estimation equation is robust.

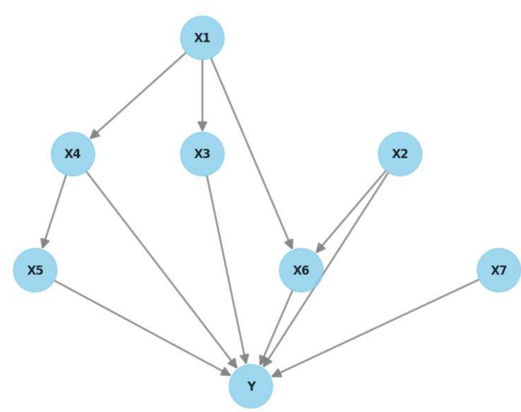

Figure 1. Directed Acyclic Graph of Delayed Ordered Material Arrival

III. Result and Discussion

Directed Acyclic Graph (DAG)

A DAG that visualises the influence between factors that may affect material arrival delays is shown in Figure 1.

In Figure 1, variable X1 is the material, variable X2 is the material ordering period (in months), variable X3 is the order quantity, variable X4 is the supplier, variable X5 is the distance, variable X6 is the requisition-to-order lead time, variable X7 is the delivery lead time. Y is the difference between the GR date and the delivery date, indicating the material delay if it is positive. The arrows on the DAG show the causal relationship between the origin and destination variables. For example, an arrow from X1 to X3 indicates that variable X1 affects X3. The variable chosen for treatment is variable X6 because requisition-to-order lead time is a factor that can be controlled. In contrast, variable Y is the outcome variable. Thus, the confounder variables are X1 and X2 because X1 and X2 affect the treatment and outcome variables. The variables that become effect modifiers are X3, X4, X5, and X7 because they only affect the outcome variable without affecting the treatment variable.

Causal Inference Analysis

At the identification stage, the effect of the treatment is identified with a "backdoor criterion" using the DoWhy library. Figure 2 shows the results of identifying the effects of treatment.

Based on the results of the DoWhy program, the estimand equation shown in equation 1 is obtained.

$$\frac{d}{d[X6]}(E[Y|X2, X1]) \qquad (1)$$

Equation 1 is the differentiation equation of the conditional expected value function Y given the values of X2 and X1, which are common cause

```
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
Estimand expression:
  d
─────(E[Y|X1,X2])
d[X₆]
Estimand assumption 1, Unconfoundedness: If U→{X6} and U→Y then P(Y|X6,X1,X2,U) = P(Y|X6,X1,X2)
```

Figure 2 Treatment's Effect Identification

```
## Realized estimand
b: Y~X6+X1+X2+X6*X4+X6*X5+X6*X3+X6*X7
Target units: ate

## Estimate
Mean value: -1.1178635741745067
p-value: [6.05334416e-15]
95.0% confidence interval: (-1.4289813128251339, -0.9046964214996487)
```

Figure 3 Causal Effect Estimation

variables against X6, which is the treatment variable. This estimand fulfils the following assumption.

"If U → X6 and U → Y, then P(Y | X6, X1, X2, U) = P(Y | X6, X1, X2)".

After the identification step, the estimation step is carried out using the linear regression method using the DoWhy library. Figure 3 shows the results of the treatment effect estimation.

The regression equation to estimate the value of Y is shown in Equation 2.

$$\mathbf{Y \sim X6+X2+X1+X6\times X3+X6\times X4+X6\times X5+X6\times X7} \qquad \mathbf{(2)}$$

The estimation results show that the mean value of the ATE value is equal to -1.1179. The ATE (Average Treatment Effect) value is the average outcome difference among all units if all units are assigned a treatment value and if all units are untreated. The mathematical notation of ATE is

$$\boldsymbol{ATE = E[Y^{a=1} - Y^{a=0}]}$$

The ATE calculation in Figure 3 is done by calculating the average Y value if the value of X6 = 1 minus the average Y value if the value of X6 = 0. It can also be interpreted that by increasing the value of X6 by one point (one day), the average Y value will decrease by -1.1179 days or decrease the delay in ordered material arrival by one day. Thus, if the company controls the requisition-to-order lead time variable, where the lead time is increased by one day, it is expected to decrease the average delay time by one day. The p-value in the significance test is 0.00. The hypothesis for the significance test on the linear regression equation is as follows.

H_0: There is no influence between the predictor and response variables.

H_1: There is an influence between the predictor and response variables.

The commonly used significance level is $\alpha = 0.05$. Because the p-value obtained is smaller than 0.05, it is concluded that there is a significant influence between the variables that are predictors in the linear regression equation on the response variable. The ATE value at the 95% confidence interval is -1.4233 < ATE < -0.9282.

Table 1. Refutation Test

Refutation Methods	Estimation Result Difference	p-value
Add A Random Common Cause	0.00001	0.9919
Use A Subset of Data	0.09608	0.7277

This indicates that with a 95% confidence level, the population ATE value will be from -1.4233 to 0.9282.

After the estimation step, the refutation step is carried out. Table 1 shows the results of the refutation test with the method of adding a random common cause and using a subset of data.

Based on the refutation test results in Table 1, the estimation results are not significantly different from the initial and p-value > 0.05. Thus, the estimation results obtained are robust to changes if synthetic independent random variables are added to the original data and if the estimation is done by taking a random subset of data.

IV. CONCLUSION

This paper concludes that the factors that affect the delay in material arrival are the material ordered by the company, the material order period (in months), the material order quantity, the supplier, the distance from the supplier to the company, the requisition-to-order lead time, and the supplier lead time. The factor that can be intervened is requisition-to-order lead time. By intervening or controlling the requisition-to-order lead time factor by one day, there is a decrease in the average time delay in material arrival by one day.

REFERENCES

Arif, S., & MacNeil, M. A. (2023). Applying the structural causal model framework for observational causal inference in ecology. *Ecological Monographs, 93*(1). doi:10.1002/ecm.1554

Chen, X., Abualdenien, J., Singh, M. M., Borrmann, A., & Geyer, P. (2022). Introducing causal inference in the energy-efficient building design. *Energy and Buildings, 277*, 1-14. doi:10.1016/j.enbuild.2022.112583

Colicchia, C., Dallari, F., & Melacini, M. (2010). Increasing supply chain resilience in a global sourcing context. *Production Planning and Control, 21*(7), 680-694. doi:10.1080/09537280903551969

Dolgui, A., Ben-Ammar, O., Hnainen, F., & Louly, M. A. (2013). A State of the Art on Supply Planning and Inventory Control under Lead Time Uncertainty. *Studies in Informatics and Control, 2*, 255-268.

Feng, Y., Wang, Y., Wu, L., Wu, L., Shu, Q., Li, H., & Yang, X. (2023). Causal relationship between outdoor atmospheric quality and pediatric asthma visits in hangzhou. *Heliyon, 9*(3).

Heizer, J., Render, B., & Munson, C. (2020). *Operations Management: Sustainability and Supply Chain Management*. Pearson Education Limited.

Kaban, M. G., & Wicaksono, P. A. (2020). Analisis Dan Mitigasi Risiko Rantai Pasok Pada Pengadaan Material Produksi Dengan Model House Of Risk (HOR) Pada Industri Mebel (Studi Kasus PT. XYZ). *Industrial Engineering Online Journal, 9*(1). Retrieved from https://ejournal3.undip.ac.id/index.php/ieoj/article/view/32987

Oprescu, M., Syrgkanis, V., Battocchi, K., Hei, M., & Lewis, G. (2019). *EconML: A Machine Learning Library for Estimating Heterogeneous Treatment Effects*.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge: Cambridge University Press.

Pujawan, I. N., & Mahendrawati, E. (2017). *Supply Chain Management*. Yogyakarta: Andi.

Putra, R. R., & Vikaliana, R. (2022). Pengaruh Defect dan Lead Time pada Lini Distribusi di PT Lasindo Jaya Bersama. *Jurnal Abiwara*, 116-124.

Sharma, A., & Kiciman, E. (2020). . DoWhy: An End-to-End Library for Causal Inference. doi:10.48550/arXiv.2011.04216

Sprenger, J., & Weinberger, N. (2021). Simpson's Paradox. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson

Sun, Q., Zheng, T., Zheng, X., Cao, M., Zhang, B., & Jiang, S. (2023). Causal interpretation for groundwater exploitation strategy in a coastal aquifer. *Science of the Total Environment*, 867. doi:10.1016/j.scitotenv.2023.161443