**Disclosure plan**

* We agree with immediate disclosure of code and data with publication.

## Data description (experiment)

## Dataset "df\_test.csv":

## Dataset description: a subset of data (3000 SKUs) that used for testing the various inventory replenishment model performance in the offline experiment.

## Dataset dictionary:

## “SKU”: disguised SKU ID

## “demand\_mean”: average demand of the SKU

## “demand\_std”: standard deviation of demand of the SKU

## “review\_period”: item review period

## “vendor\_vlt\_mean”: average vendor lead time

## “vendor\_vlt\_std”: standard deviation of vendor lead time

## “initial\_stock”: initial inventory level

## “demand\_hist”: historical demand

## “OPT\_pred”: offline optimal solution given the realized demand and vlt in the test period

## “E2E\_RNN\_pred”: replenishment quantity predicted by the proposed E2E model

## “gbm\_pred\_pred”: replenishment quantity predicted by the LightGBM model

## “sf\_rnn”: demand forecasting series outputted by the proposed E2E model

## “vlt”: vlt forecast outputted by the proposed E2E model

## The original dataset has the same dictionaries.

## Disguise process: desensitization of the SKU ID. We also randomly sample a subset of 3000 SKU from the original dataset because the original dataset is too large to disclose.

## Datasets “order\_e2e\_post\_500.csv”, “order\_e2e\_post\_500.csv”, “order\_e2e\_post\_500.csv” and “order\_e2e\_post\_500.csv”.

## Dataset description: “order\_e2e\_post\_500.csv” and “order\_e2e\_post\_500.csv” include post- and pre- experiment results, respectively, for a subset of the treatment group (500 SKU-DC pairs). “order\_e2e\_post\_500.csv” and “order\_e2e\_post\_500.csv” include post- and pre- experiment results, respectively, for a subset (500 SKU-DC pairs) of the control group. Note that the control group has been selected from all candidate SKU-DC pairs in the same category by propensity score matching.

## Data dictionary:

## “sku\_dc\_pair\_index”: desensitized index for each of the (SKU, DC) pair

## “complete\_tm”: order complete time

## “order\_amount”: order amount (replenishment quantities)

## “test\_inv”: inventory curve within test period

## “test\_demand”: demand observations within test period

## “ave\_demand”: average demand

## “ave\_inv”: average inventory level

## “vlt”: vendor lead time

## “vlt\_num”: extracted value of vendor lead time

## Dataset dictionaries of original datasets:

## Two columns “SKU\_id”: SKU ID and “DC”: distribution center ID instead of “sku\_dc\_pair\_index”.

## Other columns are the same as the disguised dataset

## Disguise process: The sku\_dc\_pair\_index has been desensitized to disguise the original SKU name and DC ID. We randomly sampled 500 (SKU,DC) pairs for the treatment group because the original dataset is too large to be disclosed.

## Dataset “control\_demand\_vlt.csv”：

## Dataset discerption: These (SKU, DC) pairs are all candidate pairs that can be selected as the control group. The propensity score matching are performed on this dataset to select a control group to reduce the potential selection bias.

## Dataset dictionary:

## Index: (SKU, DC) pair index

## “ave\_demand”: average demand

## “vlt\_num”: vendor lead time value

## Dataset dictionaries of original datasets:

## Two additional columns “SKU\_id”: SKU ID and “DC”: distribution center ID

## Other columns are the same as the disguised dataset

## Disguise process: we desensitized the original SKU name and DC ID and give each (SKU, DC) pair an index.

## Data description (model)

## Dataset "rdc\_sales\_1320.csv":

## Dataset description: a disguised example data set (5 SKU-DC pairs) for daily sales.

## Dataset dictionary:

## “row”: disguised SKU-DC ID (format: SKU\_ID#DC\_ID)

## Other than “row”, the following columns represents daily sales data. The index of each column was converted to date-time format in “data\_parser.py”.

## The original dataset has the same dictionaries.

## Disguise process: desensitization of the SKU ID and DC ID. We also randomly sample a subset of 5 SKU from the original dataset to give an example.

## Dataset “stock.csv”:

## Dataset description: “stock.csv” is a disguised example data set for stock levels in the warehouse.

## Data dictionary:

## “item\_sku\_id”: disguised SKU-DC ID (format: SKU\_ID#DC\_ID)

## Other than “row”, the following columns represents daily stock-level data. Each column is indexed with date.

## The original dataset includes longer record of stock level. The provided data set is trimmed to one month.

## Disguise process: desensitization of the SKU ID and DC ID. 2 SKUs are randomly picked from the original dataset and is trimmed to be one-month long.

## Dataset “vlt\_2019\_0305.csv”：

## Dataset discerption: a disguised example dataset for vendor lead time.

## Dataset dictionary:

## “item\_sku\_id”: disguised SKU- ID

## “int\_org\_num”: disguised DC-ID

## “pur\_bill\_id”: purchase bill ID.

## “create\_tm” and “complete\_tm”: create time and complete time.

## “vlt\_actual”: actual vendor lead time (equals to complete time – create time).

## “actual\_pur\_qtty”: actual purchase quantity.

## “item\_first\_cate\_cd”, “item\_second\_cate\_cd”, “item\_third\_cate\_cd”: item first, second, and third category ID.

## “brand\_code”: brand code.

## “create\_day\_of\_week”: create day of the week.

## Disguised volumn names for three groups of features: item dimension information (e.g. weight, height, length), item-wise and vendor-wise historical VLT statistics (e.g. averaged VLT over certain time window, standard deviation of VLT ), statistics of item historical purchase quantities (e.g. averaged purchase quantity over certain time windows).

## Dataset dictionaries of original datasets:

## More column corresponding to item dimension information, item-wise and vendor-wise historical VLT statistics. The column names should be consistent with “VLT\_FEA” defined in config/config.py

## Other columns are the same as the disguised dataset

## Disguise process: we desensitized the original SKU name and DC ID and remove confidential column names. The data set contains on SKU-DC pair as an example.

**Code (experiment)**

1. “E2E\_Section4.ipynb”: reproduce Figure 2

* Same code applies on the disguised and undisguised data.
* Description: Offline comparisons of seven different inventory replenishment models. As only a subset of the entire test dataset is allowed to share due to business reasons. The reproduced results have different numerical value compared to the results in Section 4.2 that are obtained on the entire test dataset. However, all the observed trend, relative performance comparison and conclusion remain the same.

1. “propensity\_score\_matching.ipynb”: reproduce Figure 3

* Same code applies on the disguised and undisguised data.
* Description: Demonstration of using propensity score matching to select the control group out of all candidate SKU-DC pairs. As we only provide a subset of the treatment group, the propensity score density plots are different than plots in the paper. But it still has a perfect matching.

1. “field\_experiment\_summary.ipynb”: reproduce Table 2,3 and 4

* Same code applies on the disguised and undisguised data
* Description: evaluates five inventory performance metrics for both the treatment and control groups for the pre-experiment periods and post-experiment periods. This notebook also demonstrates the result of the t-test, linear regression and Difference-in-Difference estimation. As we only can share a subset of the dataset, the numerical value is different from the results in Section 5 but the conclusion remains the same.

**Code (model)**

The folder “E2E-model-code” includes code and pseudo-code for constructing and training the End-to-End model with RNN.

Due to confidentiality, certain parts of the source code are replaced with pseudo-code or detailed description.

Disguised parts:

1. “config/config.py”: Details of feature names and RNN parameters are disguised.
2. “data\_loader/data\_parser.py”: Disguised input and output paths. Disguised value of some parameters. Disguised data cleaning and preprocessing details in functions “add\_target” and “dummy\_cut”.
3. “models/loss.py”: Disguised weight parameters in training loss functions. Disguised details of choice of loss function.
4. “models/model.py”: Disguised initialization of weights. Disguised feature dimensions.
5. “utils/demand\_pkg.py”: Window lists used in features are disguised.