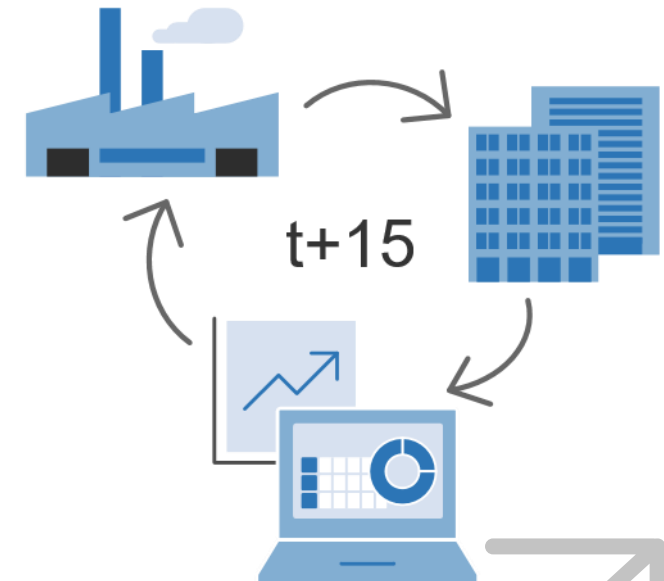# Experimental Short-Term Statistics based on

Data Imputation Methods

t+15

Our Project Goal

We are improving and accelerating the supply of short-term economic indicators in Germany!

The aim of our project is to improve processes and integrate them into the compilation of statistics to provide additional key economic figures much faster than before:

→ Approximately at time t+15 (currently t+30 to t+70).

Firm-level data based method: Imputation

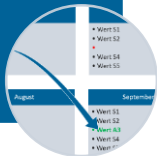## Our tested imputation methods

- LOCF
- Mean

- Random Forest
- Cart

- Predictive Mean Matching
- Weighted Predictive Mean Matching
- Random-hot-deck

- Bayesian linear regression
- Linear regression ignoring model error
- Lasso linear regression
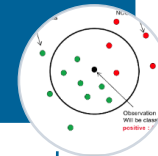- Linear regression, predicted values

Univariate imputation

Tree based methods

Matching methods

Regression based methods

These methods simply use the character (i.o.c. turnover) itself
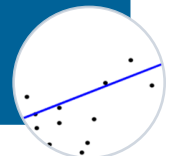
Decisions are made at different nodes within these models

The information needed is taken from another (similar) unit

Linear relationships are used to estimate the target variable

Measures of determination for the variables used to estimate the new orders index

	Domestic incoming orders	Foreign incoming orders	Foreign incoming orders (non-EU)
Type of Unit	0,0106	0,0083	0,0057
Federal state	0,0026	0,0017	0,0013
Industry 4-digit	0,0835	0,0796	0,0690
Turnover (domestic/foreign/non-EU) of the previous month	0,6192	0,3785	0,6644
Incoming orders (domestic/foreign/non-EU) of the previous month	0,6095	0,8021	0,7727
employees of the previous month	0,4614	0,5388	0,4613

Our model estimates the current period new orders for company for a company indexed by *i* by the following variables:

$$\widehat{Incoming\ oders}_{t,i} = \hat{\beta}_0 + \textbf{Federal state}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_1 + \textbf{Type of unit}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_2 + \textbf{Industry} - \textbf{4} - \textbf{digit}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_3 + \textbf{Reporting month}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_4$$

$$+ \textbf{Turnover class}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_5 + \textbf{New orders class}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_6 + \textbf{Timely notification}_{t,i}^{T}\ \widehat{\boldsymbol{\beta}}_7 + \text{Employees}_{t-1,i}\ \hat{\beta}_8 + \text{Rate of Chance}_t\ \hat{\beta}_9$$

Variable	Description
New orders category of the previous month	Previous month' new orders has a high informative value about the new orders of the current month
Turnover category of the previous month	Previous month' turnover has a high informative value about the new orders of the current month
Employees of the previous month	Have informative value about the size of the company and thus the new orders of the company
Federal State, type of business, Economic sector (WZ 4 digits)	These variables do not change over time. Therefore, the previous month' information can be used
Rate of change of WZ 2 digits (German industrial classification) in the reporting month	The average rate of change takes into account the (cyclical) development in the WZ as of the current reporting period.
Time period variable reflecting the development of turnover in the estimated month	This variable reflects time fixed effects. Hence, this variable can be used to estimate a coefficient for the effect of an observation being in a particular reporting period.

Measures of determination for the variables used to estimate the turnover index

	Domestic turnover	Foreign turnover	Foreign turnover (non-EU)
Type of Unit	0,0088	0,0073	0,0051
Federal state	0,0049	0,0017	0,0011
Industry 4-digit	0,1215	0,1106	0,0981
Turnover (domestic/foreign/non-EU) of the previous month	0,8992	0,8975	0,8908
Incoming orders (domestic/foreign/non-EU) of the previous month	0,3241	0,7038	0,6798
employees of the previous month	0,2837	0,5915	0,5190

Our model estimates the current period turnover for company a company indexed by *i* by the following variables:

$$\widehat{Turnover}_{t,i} = \hat{\beta}_0 + \textbf{Federal State}_{t,i}^{T}\,\hat{\beta}_1 + \textbf{Type of Business}_{t,i}^{T}\,\hat{\beta}_2 + \textbf{WZ} - \textbf{4} - \textbf{digits}_{t,i}^{T}\,\hat{\beta}_3$$

$$+ \textbf{Reporting period}_{t,i}^{T}\,\hat{\beta}_4 + \textbf{Turnover category}_{t-1,i}^{T}\,\hat{\beta}_5 + Employees_{t-1,i}\,\hat{\beta}_6 + Average\ change\ rate_{t}\,\hat{\beta}_7$$

Variable	Description
Turnover category of the previous month	Previous month' *turnover* has a high informative value about the turnover of the current month
Employees of the previous month	Have informative value about the size of the company and thus the turnover of the company
Federal State, type of business, Economic sector (WZ 4 digits)	These variables do not change over time. Hence, this information from previous months can be used
Rate of change of WZ 2 digits (German industrial classification) in the reporting month	The average rate of change takes into account the (cyclical) development in the WZ as of the current reporting period.
Time period variable reflecting the development of turnover in the estimated month	This variable reflects time fixed effects. Thus, this variable can be used to estimate a coefficient for the effect of an observation being in a particular reporting period.

Comparison of separate models per industry and global models across all industries

| New orders index	MAE	RMSE
Global model	2.737	3.320
Speparate models	2.339	3.105

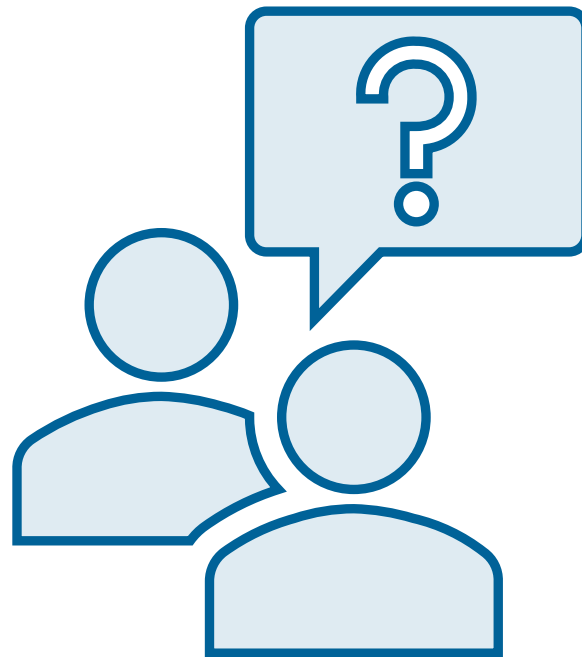| Turnover index	MAE	RMSE
Global model	2.158	2.573
Speparate models	1.879	2.306
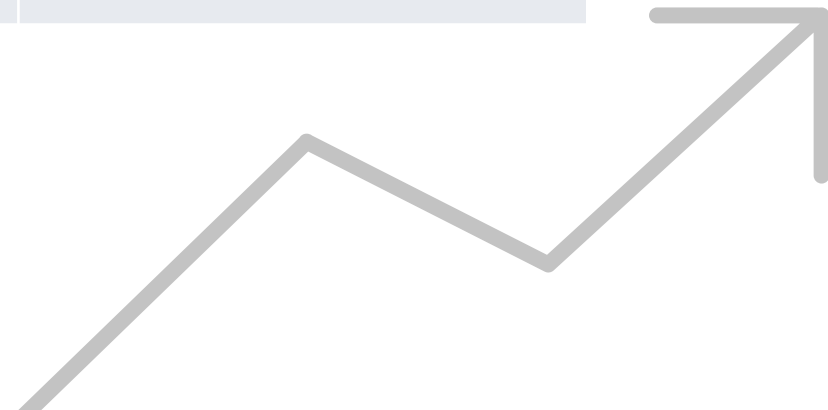
Problems

- Only a few regressors reflect the economic development
- Strong fluctuation in *order intake* due to major orders
- Due to a lack of alternatives, regressors had to be included in the model that actually do not have a high explanatory power
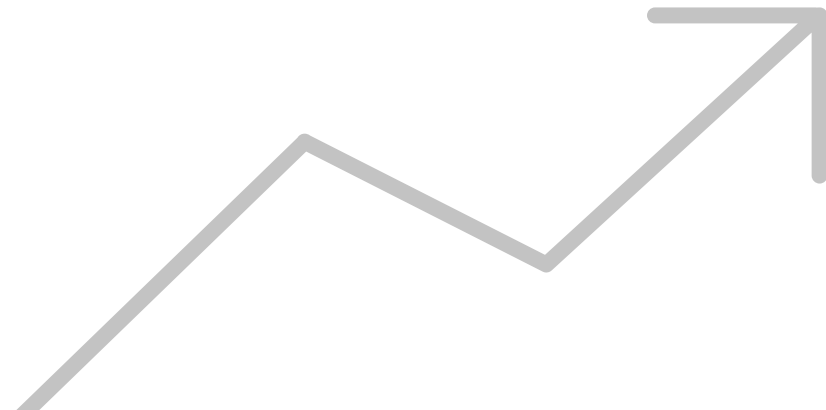
Are there any questions?

Thank you very much for your attention!

Jan Ditscheid
Jan.ditscheid@destatis.de
Phone: +49 611 75 3980