NSM 2022
NORDIC STATISTICAL MEETING

Using electricity data to identify construction activity

Simon B. Halifax, Statistics Denmark, sbh@dst.dk

Abstract

Measuring activity in the construction sector is an important indicator for use as short-term business statistic. We have good registry data on permitted and completed projects, but data about when the construction actually begins are usually not available before completion. A model to correct for the delay is our current setup allowing for a quarterly statistic on the number of started construction projects for every month. The accuracy of the model is limited and the data published 5 weeks after the end of a quarter.

Therefore, there is a potential for faster and more accurate statistics. With our new access to smart-meters with high frequent data on electricity consumption for every address in Denmark this potential can be released. We receive information on electricity consumption agreements once every month only 8 days after the end of the reference month. So if we can connect the advent of a new agreement and consumption of electricity on an address with information on permitted construction projects we are able to make a fast indicator of construction activity.

A pilot study was conducted identifying a single construction using consumption data and it proved to be much more reliable than the registry data. The registry data that had an incorrect date of construction start and was not available before construction end 4 months late. Going forward the idea is now to identify whether this can be extended to more construction projects and compare the results with our current quarterly statistic.

Initially, data will be made available as an experimental indicator. If it proves to be stable and reliable, the hope is to support or replace the current official statistic with a faster and potentially better one.

Keywords: constriction activity, smart-meters, electricity consumption, fast indicators, experimental statistics.

Introduction

Short-term business statistics is all about indicators being available as fast as possible to assess the immediate state of the economy. One such indicator is the activity in the construction sector as measured by the number of commenced projects during a certain period. At Statistics Denmark our current indicator is based on registry data where monthly data for a given quarter is published with a delay of 5 weeks. These figures have proven to be deficient as data about commenced projects are usually not reported until the project is complete. The total number of commenced projects during a quarter is not fully available until ~18 months after the end of the quarter. We have implemented a statistical machine learning model finding tendencies in the delay to correct for this late registry of commenced project improving the accuracy significantly. There is, however, still much room for additional improvements to reduce the delay of the indicator.

Because of the nationwide change to smart-meters for electricity consumption we have incoming data for all metering points with an interval between 15 minutes and one hour. This high-frequent data is available with a delay of 8 days allowing us early to know where and how much electricity is consumed. The meters are linked uniquely to an address as well as an agreement with information about the consumer, but the latter is of limited concern for this study.

We can get an early indication of whether a construction project has been commenced by combining this data with knowledge about permitted construction projects on address level. This is particularly useful regarding construction of single houses on empty plots as these are usually unambiguous in both data sets.

Pilot study

To test our hypothesis we looked into a specific project that we were able to follow from the purchase of an empty plot to the completion of the building, which in this case was a summer home. The table below show the dates of the notices we get regarding the project in our registry of construction projects and electricity data.

Event	Time of event	Available in registry	Quality
Permit of project	March 21st	April 24th	Good
Building start	March 31st	October 24th	Wrong date

Finished project	July 4th	October 24th	Wrong date
Hand-over	October 10th	October 24th	Correct
First use of electricity	July 18th	After 8 days	Good

The first time the project appears, is when the permit is registered. This data is available in our quality assured registry ~3 weeks after the end of a given month where we receive an updated status of all projects and therefore includes newly added projects that have been permitted since last update. In our registry, nothing happens for the next 6 months until the update in October where data about both start up and completion of the project appears. For this specific project, we know that the data is wrong because we follow the actual construction. Generally we do not know this and accept the data we get from the municipalities when they open and close construction projects. This implicates that the quality of our current statistic is not as expected and thus making it harder to use the statistic to assess economic development and qualify policymaking.

In our official statistics the project appears as a permitted one in our first publication of figures for the first quarter of the year as it was permitted in march. When the data for the fourth quarter of the year is published the data for first and third quarter are revised because the new information from the update in October now counts the project as commenced in the first quarter and finished in the third quarter.

Information about commenced projects in the first quarter of a year in February the next year is of limited value for short-term statistics so in order to improve this we looked at the electricity data. Looking into microdata, we see that a new electricity agreement started in June and that consumption started on July 18th. This is the actual date the project commenced and we can follow the daily (or even hourly) consumption pattern throughout the coming period. This pattern is also of interest in its own right as it might be extended for finding other types of construction projects that does not follow the simplicity of this one with a single house on an empty plot.

Looking on the pattern at year-end we clearly see the hand-over as the consumption now show a weekend based use of the new summer home.

Process for generalization

The pilot showed us that we can speed up the knowledge of commenced construction projects of single buildings on empty plots with several months by using electricity data rather than looking on the registry alone. It is however not viable to look at microdata for this use so we have to make a setup where we can utilize the information in the electricity data without looking into hourly consumption data for each address linked to a project as this is a very big dataset.

With this in mind we chose to look at all permits given from 2019 and ahead regarding construction of new single houses, townhouses, and summer homes. Most of these projects have already been finished, but we can see in our registry what information was available at any given monthly update. For electricity data we convert the high-frequency data to monthly data to ease the combination of the data sets. With the data in hand the idea is now to calculate the delay between the permit and the moment that information about when the project was commenced is available. Note that this is not the time it takes from a project is permitted to the construction starts. Then we will try to see if we can uniquely identify a metering point. Here it turns out that many addresses have several metering points. For those with only one metering point we would like to see a pattern of very limited or ideally no electricity consumption prior to and shortly after the permit. Sometime after the permit we would like to see an increase in electricity consumption indicating that the construction project has commenced.

Having identified an increase in electricity consumption from a low level after the permit of a project we now compare the timing with what we get from the registry i.e. do electricity data tell us something we do not yet know or is the data already up to date?

At this point we have several options to utilize this data. We could fill in the registry with a proxy variable for the commencing date with the month where we observe consumption of electricity. This might be an even better indicator for measuring the actual activity in the building sector because we know that the values filled in the records for construction start mainly have statistical value and the municipalities have little interest in ensuring that the dates are correct. We use the data for (somewhat) short-term statistics as it is an important indicator for the state of the economy whether more or less construction projects are started in a given month.

Another use of this data is to use it as an indicator in its own right. Since we might have a good indicator for a certain subgroup of the entire activity in the construction sector this might also be valuable for indicating the trend of the entire sector. Where our current statistic covers all kinds of construction projects concerning type of building, owner, type of project etc. the quick new indicator focuses only on the construction on new single houses, townhouses or summer homes with only one smart-meter on the address. To assess the quality and reliability of this indicator we need to analyse how much of the volume of construction projects is covered by this subgroup. The same comparison are to be done on our existing statistic to see if the fraction appears somewhat constant. If it does it might allow us to extend the trend beyond the smaller subgroup.

The sample space of construction projects with electricity consumption on the address added are quite huge so the following list show how projects are selected to be a part of the exploratory analysis of the new indicator.

- The registry for construction permits is the base dataset with these filters:
    - New buildings
    - Type of buildings include single family homes, summer homes and townhouses.
    - Permit date is in year 2019-2021
- Monthly electricity consumption for the unique address in the project with these filters
    - Only addresses with a single metering point are included
    - Consumption data starting in the month before permit and including the following 14 months.

These filters are all subject to change and adjustments depending on the results of the analysis. For the electricity data the long period is used to identify patterns and especially to find different outcomes that all fit into the purpose of the indicator. As many newly constructed buildings are replacements of existing buildings the ideal of a zero consumption prior to construction start is only true for a limited number of projects. The idea is to include as many projects as possible without having data that distorts the general picture or makes it hard to find patterns and tendencies.

When looking at the raw numbers of permitted new constructions the table below show what is removed every time we make a restriction to simplify the dataset to make the indicator fast and viable.

Results

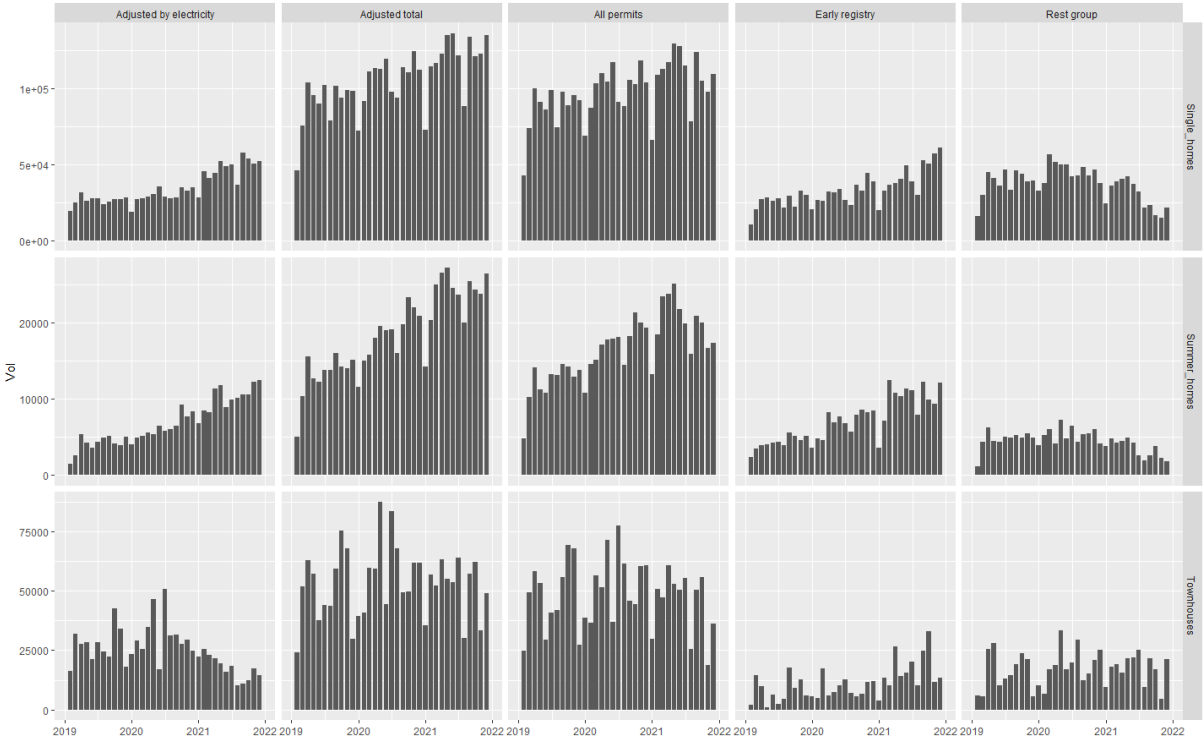
As mentioned above the total dataset of permitted project is reduced as restrictions regarding the availability and usability of electricity consumption data is added. Looking at the permitted projects from 2019-2021 it turns out at we can identify quite a large fraction with a uniquely identifiable smart-meter (MPO). Of this subset we remove the projects that already have a date of commencing. It turns out that around one third of the construction projects is now left in our dataset where we can use electricity data to estimate the unknown date of commencing.

	Fraction with unique MPO	Fraction with delayed start and consumption
Single homes	~65%	~30%
Summer homes	~75%	~30%
Townhouses	~80%	~45%

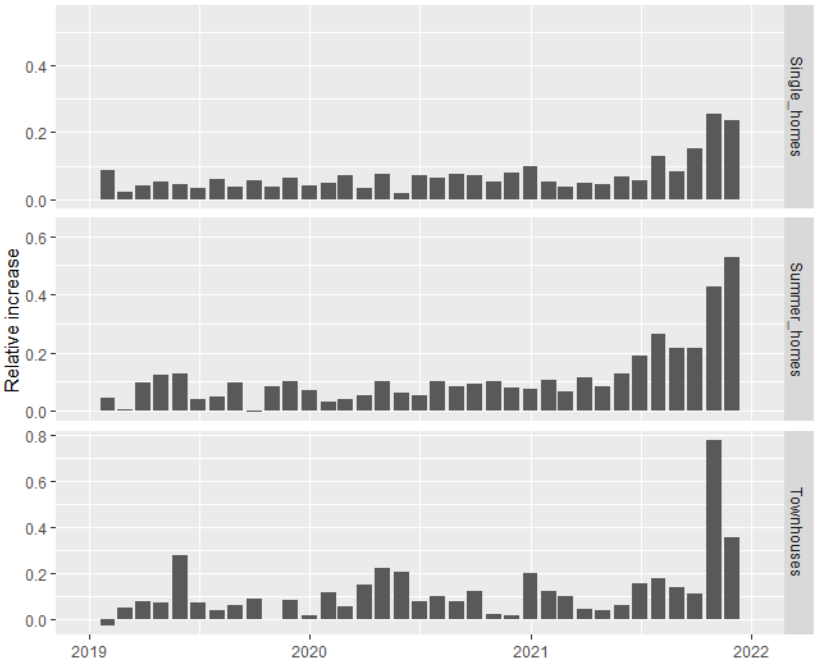
The plot below show the result of the analysis split into the three types of buildings. The first column show the monthly total square meters of all the projects that is estimated by electricity data as defined above. The date is either the registry date of commencing of the project if it is available or the first month after permit where the consumption of electricity exceeds 20 kWh. So to get an idea of the potential the second and third column are to be compared as the third shows what is known to us from the registry and the second have been adjusted with the results from the first column. The forth column contains the projects where there is no delay in the date of commencing as it is noted along with the permit. This fifth column show the remaining projects which mainly consists of those that we cannot identify a single metering point. The second column is the sum of the first, the fourth and the fifth and is included is it therefore can be compared with the raw permits seen in column 3.

It is visually clear that we see a tendency in the adjusted data towards higher activity in the end of the period consistent with what we know from other sources of the state

of the construction sector. Below is another plot showing the relative changes what the adjustment brings.



The actual figures are hard to read on the plot above so a plot showing what fraction is added on a monthly basis for the adjusted data as compared to the raw. In the analysis the newest available electricity data is from January 2022 so the plot is restricted to that period. This is due to be improved in nearest future allowing for an even timelier indicator. It is clear that this new approach can correct for quite a lot of the delay with double-digit percentage increases for the last six months.



Further work

As the plots show we have an indicator now that might show an economic tendency and we will study the newest data when it becomes available. This will allow us decide whether the indicator is strong enough to use as a fast index of construction activity.

The process now goes on to implementing this solution and comparing it to the existing one. Here we might be able to see a pattern in the amount of construction projects with missing registry data that we can identify using electricity data and then extend the current machine learning model to take this into account. Hopes are that it is even more reliable an estimator than the attempt to estimate the pattern of delay in the municipalities.

Our first priority is to make the simple and straightforward indicator and then have the opportunity to extend with the advanced implementation of updating the estimation model. As the potential of the indicator proved to be promising, using it as a measure for commenced projects is following the steps above. This means that we for each month will count the number of single building projects with a single metering point having new electricity consumption. This gives us an indicator with a delay of only 8-10 days, depending on implementation. Since we can make it monthly we also quickly get a time series that we can compare back in time to the registry-based

statistic which we at this point is considered an accurate statistic when looking at data for 2020 and earlier. This is used to validate new and incoming data on both indicators. If the experimental indicator shows the same tendencies as the current statistic we expect to be able to use it as a short-term statistic measuring construction activity.

We do note that the construction sector covers much more than new houses and summer homes e.g. public works like hospitals and universities but as our shot-term statistics get much political attention, it might help decision making on an earlier stage. As it is now it takes 1-2 years before we have reliable statistics about the activities in the construction sector from our project registry a recession would appear much faster in other statistics like for instance the unemployment. If our indicator proves reliable we would know almost instantly when the activity in the sector starts to drop and can thereby help to improve political decision making to counter an economic downfall.

Another advanced use of the electricity data is to dive even deeper into the specific electricity consumption to see whether construction activity has a pattern that is distinguishable from household consumption. This analysis could result in an ability to identify other construction projects like extensions and demolition. On that point we intend to identify the projects and then supply the microdata to our data science lab for advanced analytics and machine learning.

This study is just one in a suite of projects we are looking into here at Statistics Denmark to utilize the electricity data we have available. Other uses include the ability to identify empty dwellings, forecast enterprise bankruptcy following a drop in energy consumption and a general monitoring of economic activity across sectors.