Correlations and clustering in wholesale electricity markets

Tianyu Cui\textsuperscript{a,\,*}, Francesco Caravelli\textsuperscript{b}, Cozmin Ududec\textsuperscript{c}

\textsuperscript{a} Science & Technology on Integrated Information System Laboratory Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{b} Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
\textsuperscript{c} Invenia Labs, 27 Parkside Place, Parkside, Cambridge CB1 1HQ, UK

HIGHLIGHTS

- We studied wholesale electricity markets.
- We introduced a new correlation based on event synchronization for spiky time series.
- A clustering benchmark based on string correlation of nodal names is proposed.
- RMT with MST achieves best performance but it is unstable over time.
- Event synchronization with MST is most suitable for its stability and accuracy.

ARTICLE INFO

Article history:
Received 11 August 2017
Received in revised form 26 October 2017
Available online 24 November 2017

Keywords:
Correlation measures
Clustering
Electricity prices

ABSTRACT

We study the structure of locational marginal prices in day-ahead and real-time wholesale electricity markets. In particular, we consider the case of two North American markets and show that the price correlations contain information on the locational structure of the grid. We study various clustering methods and introduce a type of correlation function based on event synchronization for spiky time series, and another based on string correlations of location names provided by the markets. This allows us to reconstruct aspects of the locational structure of the grid.

1. Introduction

Electricity is different from other commodities in at least three important respects: it cannot currently be stored efficiently, it flows through the electricity grid according to the laws of electromagnetism rather than directly from a producer to a consumer, and at any moment supply and demand must match almost exactly to avoid blackouts or other issues. The capacity of the transmission network also limits the amounts of power that can be injected or withdrawn at certain locations. Therefore, in order for an electricity grid to function safely and efficiently, the configuration and physical limits of the grid and the locations of generation and consumption must be taken into account when making decisions about consumption and production. To this end, many grid operators around the world have adopted a spatial and temporal pricing mechanism known as Locational Marginal Pricing (LMP)\textsuperscript{[1]}. This mechanism sets potentially different prices at hundreds or thousands of important locations (nodes) throughout an electricity grid. These prices are part of the solution to an optimal power flow (OPF) problem\textsuperscript{[2]}, and represent the increase in optimized total system costs as a result of a small increase in the amount of power consumed at a location at a specific time.

\* Corresponding author.
E-mail address: cuitianyu@iscas.ac.cn (T. Cui).

https://doi.org/10.1016/j.physa.2017.11.077
0378-4371/Published by Elsevier B.V.
In this paper, we study the properties of LMPs in two North American wholesale electricity markets: Pennsylvania–New Jersey–Maryland (PJM) [3], and Midcontinent Independent System Operator (MISO) [4]. In particular, in both of these markets there are two types of LMPs: day-ahead (forward) LMPs, and real-time (spot) LMPs [5]. The day-ahead prices are outputs of what can be thought of as a planning exercise by the grid operator, in which OPF problems are run every day for each hour of the next (target day). The real-time LMPs are then the outputs of the OPF problems that are solved in real time (usually every 5 or 15 min) to adjust the schedules output from the day-ahead runs.

Our main goal is to understand what information can be inferred from day-ahead and real-time LMPs. Specifically, we are interested in whether any underlying structure can be inferred from the correlations between prices. As discussed above, LMPs result from a highly structured and constrained optimization problem [6], which induces non-local correlations between prices. It is therefore interesting to understand how prices are correlated according to various measures, and how clusters of nodes emerge, and whether these can be understood in terms of the spatial distribution of nodes in the grid.

We will use a combination of methods introduced in the machine learning and statistical physics literatures. One interesting approach begins by first filtering the relevant information from covariance and correlation matrices. This approach has mainly been used in the financial literature, in order to remove possible noise and to focus only on the most important structures. We will have to introduce a few caveats and differences which are due to the underlying physical structure of power grids [7–9]. Given the specific spiky nature of the time series, as an alternative to the Pearson correlation, we also introduce a method commonly used for neural networks, event synchronization, to measure the synchronization between spikes at various locations. We will also test another correlation measure introduced in machine learning: the graphical lasso method. This is then used to cluster the time series. For this purpose, we use both the Minimal Spanning Tree method, the Planar Maximally Filtered Graph algorithms approach and Spectral Clustering. We then compare the results to clustering based on the node names as a proxy for node location.

2. Data structure

The dataset we will analyze has been collected by Invenia TCC. A subset of the data was used in [9], where a description of the data is also provided [10]. The time series we consider is available for each node in the PJM and MISO grids for a period of 3 years (January 1, 2012 to December 31, 2015) at hourly resolution. We focus on a subset of 1287 nodes for PJM and 2568 nodes in MISO (these are nodes which virtual participants can transact on).

The full (day-ahead or real-time) price of electricity at node \( n \) and time \( t \) is given by:

\[
LMP(n, t) = MEC(t) + MCC(n, t) + MLC(n, t),
\]

where MEC stands for Marginal Energy Cost, MCC stands for Marginal Congestion Cost and MLC stands for Marginal Loss Cost. The \( MCC(n, t) \) component is the price due to transmission congestion, i.e., it is the marginal cost of supplying the next increment of load at a location, taking into account the transmission constraints of the grid. This can be positive or negative, and is often 0. For example, when a power line at some location is at its limit for carrying power, the load at that location must be serviced through another line, which can be more costly. \( MLC(n, t) \) is the price due transmission losses on the grid. This is generally small compared to \( MEC \). \( MEC(t) \) can roughly be thought of as the price of electricity at any given node if there is no congestion and loss to that node. The \( MEC \) component is independent of node, and thus represents a price shift for the whole market. The \( MLC \) and \( MCC \) components are instead time and node dependent, but we observe that in general \( MCC \gg MLC \).

Thus, for the present paper we study only the MCC component of the prices in the day-ahead market, which directly address the inefficiency of power transmission and is the main source of volatility in the LMP time series. An important quantity for many market participants and for the grid operator is the difference between day-ahead and real-time prices:

\[
\Delta(n, t) = DA(n, t) − RT(n, t).
\]

This is the time series we will focus on.

One important detail which will be important for our analysis is the fact that each nodal price has a string associated to it, which is of the form “NODENAME_CODE”. The text “NODENAME” can be loosely associated with the location of the node. We can in fact run a clustering algorithm based on the name and compare to the results based on price correlations only. This will allow us to compare correlation matrices obtained from independent methods and identify the best price correlation technique to pair nodes and associated clusters.

3. Correlations, synchronization measures, and filtering processes

We begin by introducing several measures of interdependence between time series which will be relevant in our analysis.

3.1. Pearson correlation

The Pearson correlation is one of the most commonly used and simplest measures. It assumes stationarity and a linear relationship between the time series. The correlation matrix is defined as:

\[
C_{ij} = \text{Corr}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i] \text{Var}[X_j]}},
\]

where

\[
\text{Corr}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i] \text{Var}[X_j]}},
\]

is the correlation coefficient between variables \( X_i \) and \( X_j \), \( \text{Cov}[X_i, X_j] \) is the covariance between \( X_i \) and \( X_j \), and \( \text{Var}[X_i] \) is the variance of \( X_i \).
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات