Modeling of the Electricity Spot Prices Using the Skew-Normal Distribution

The Case

- This example shows how to simulate the future behavior of electricity spot prices from a time series model fitted to historical data.
- Because electricity spot prices can exhibit large deviations, the example models the innovations using a skew-normal distribution.
- The data set in this example contains daily electricity spot prices from January 1, 2010 through November 11, 2013.

Case Overview

- The price of electricity is associated with the corresponding demand.
- Changes to population size and technological advancements suggest that electricity spot prices have a long-term trend.
- Also, given the climate of a region, the demand for electricity fluctuates according to the season.
- Consequently, electricity spot prices fluctuate similarly, suggesting the inclusion of a seasonal component in the model.

Case Overview

- During periods of high demand, spot prices can exhibit jumps when technicians supplement the current supply with power generated from less efficient methods.
- These periods of high demand suggest that the innovations distribution of the electricity spot prices is right skewed rather than symmetric.

Steps for Creation

1. Determine whether the spot price series has exponential, long-term deterministic, and seasonal trends by inspecting its time series plot and performing other statistical tests.
2. If the spot price series exhibits an exponential trend, remove it by applying the log transform to the data.
3. Assess whether the data has a long-term deterministic trend. Identify the linear component using linear regression. Identify the dominant seasonal components by applying spectral analysis to the data.

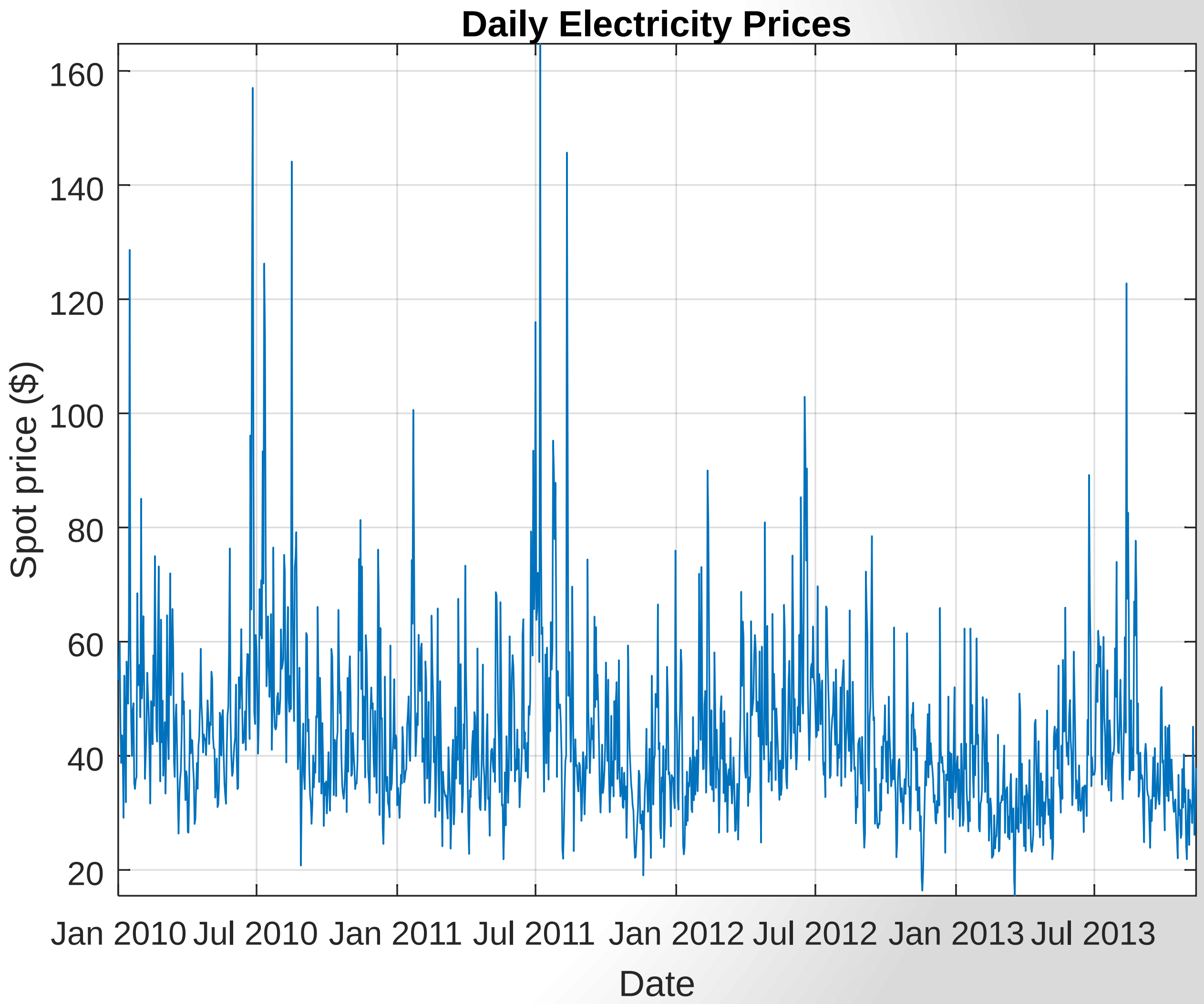
Steps for Creation

4. Perform stepwise linear regression to estimate the overall long-term deterministic trend. The resulting trend comprises linear and seasonal components with frequencies determined by the spectral analysis. Detrend the series by removing the estimated long-term deterministic trend from the data.
5. Specify and estimate an appropriate autoregressive, moving average (ARIMA) model for the detrended data by applying the Box-Jenkins methodology.
6. Fit a skew-normal probability distribution to the standardized residuals of the fitted ARIMA model. This step requires a custom probability distribution object created using the framework available in Statistics and Machine Learning Toolbox™.

Steps for Creation

7. Simulate spot prices. First, draw an iid random standardized residual series from the fitted probability distribution from step 6. Then, back-transform the simulated residuals by reversing steps 1–5.

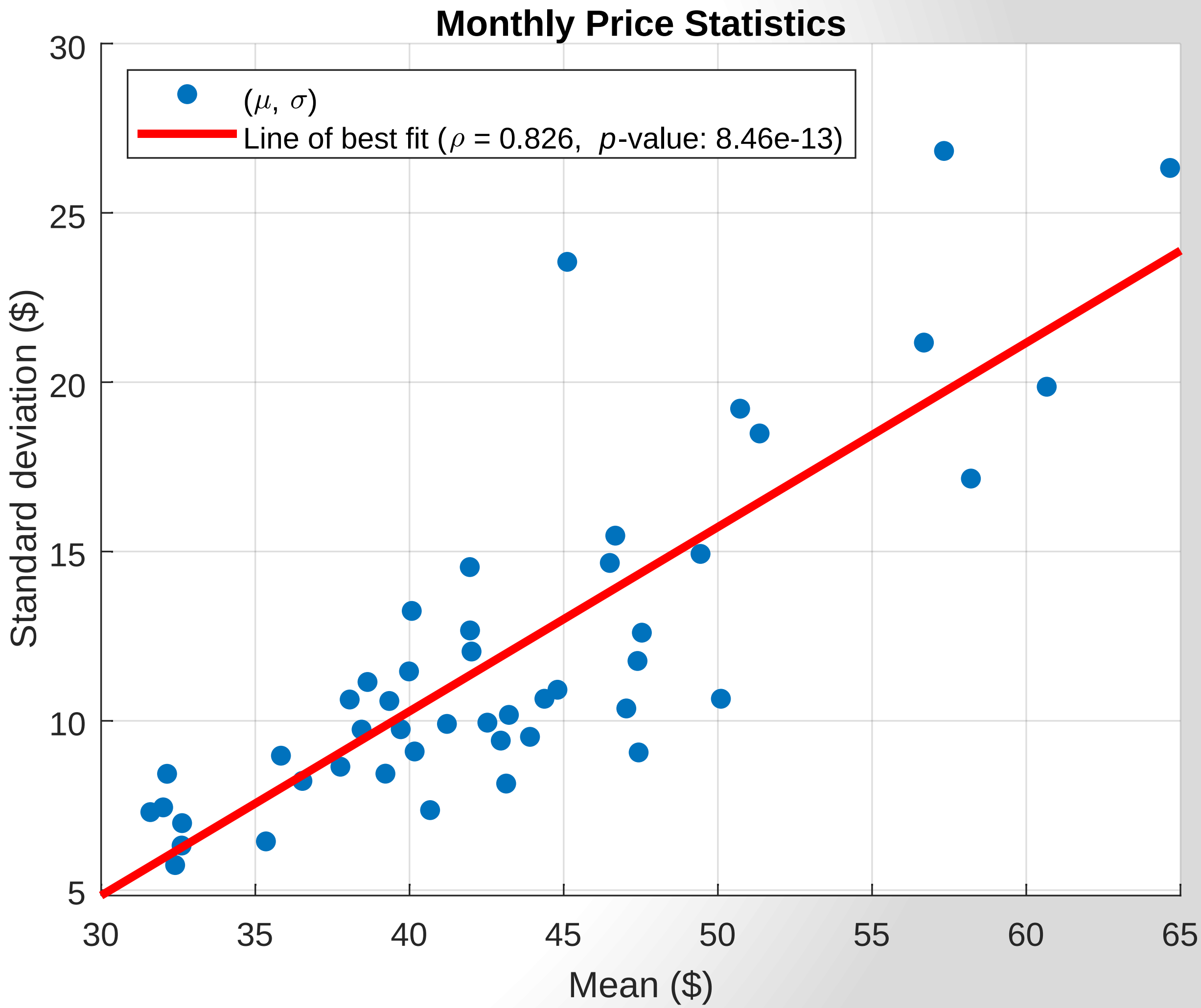
Trends Determination



Trends Determination

- The spot prices exhibit:
  - Large spikes, which are likely caused by periods of high demand
  - A seasonal pattern, which is likely caused by seasonal fluctuations in demand
  - A slight downward trend
- The presence of an exponential trend is difficult to determine from the plot.

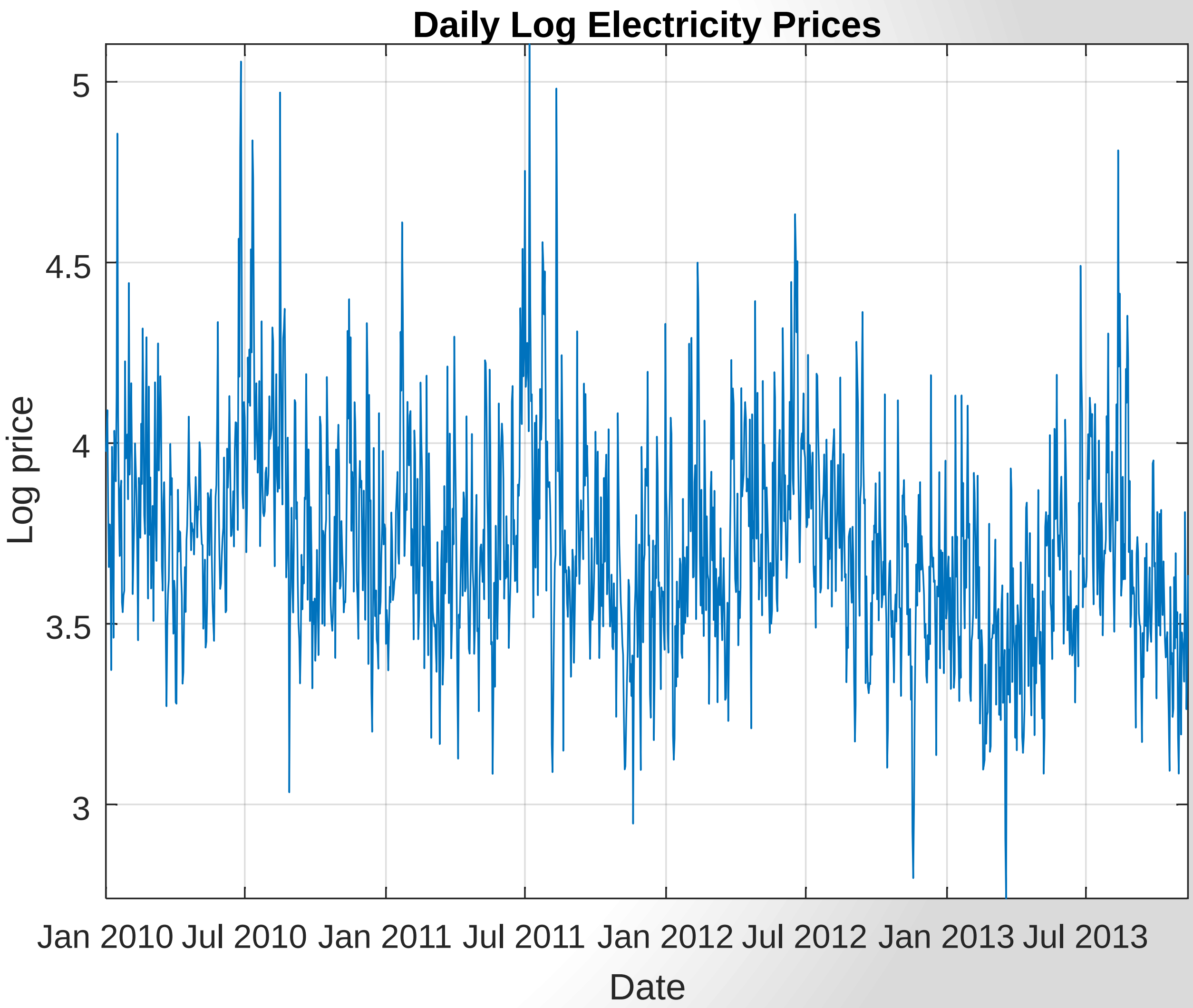
Remove Exponential Trend



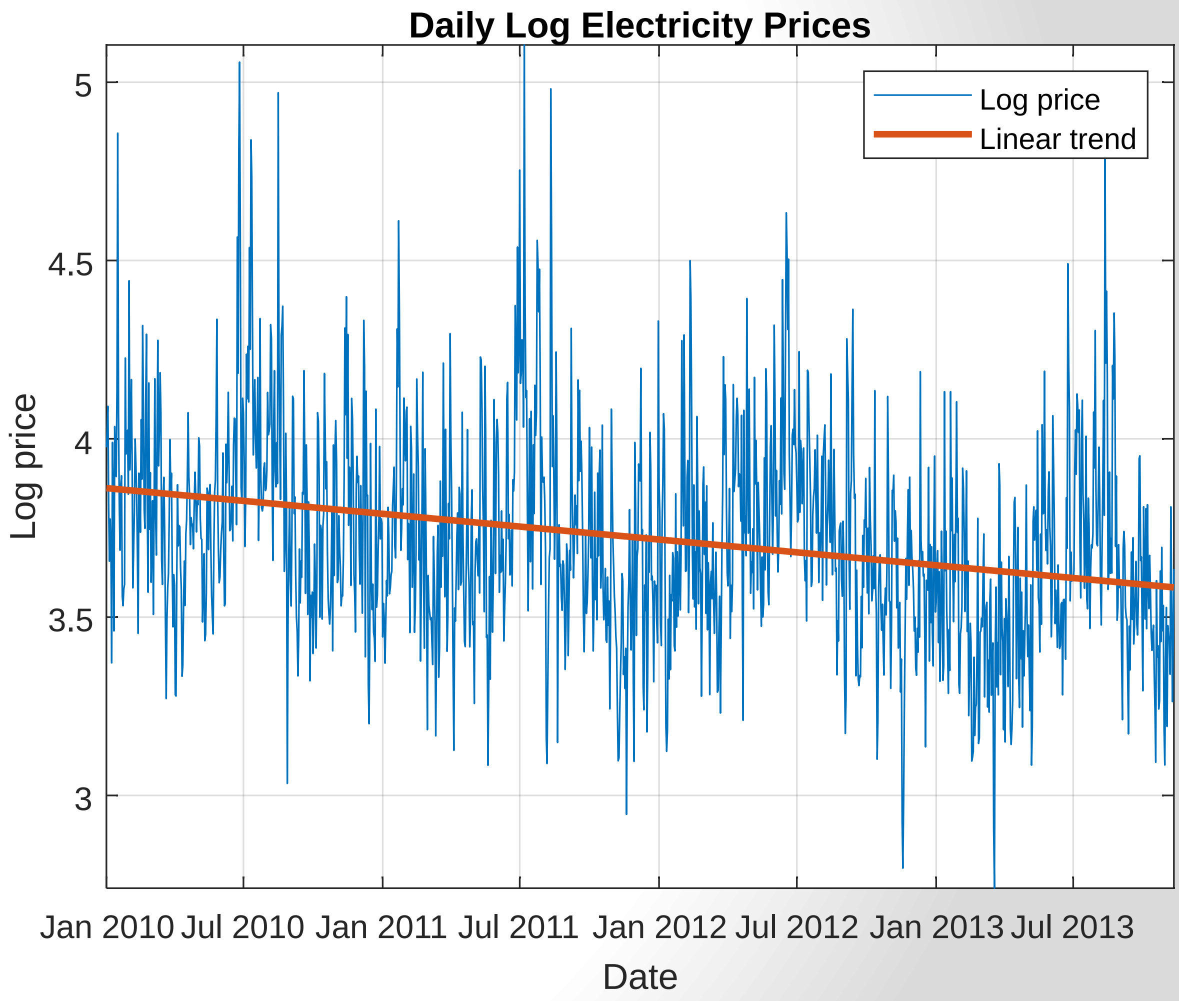
Remove Exponential Trend

- The *p*-value is close to 0, suggesting that the monthly statistics are significantly positively correlated.
- This result provides evidence of an exponential trend in the spot price series.

Remove Exponential Trend



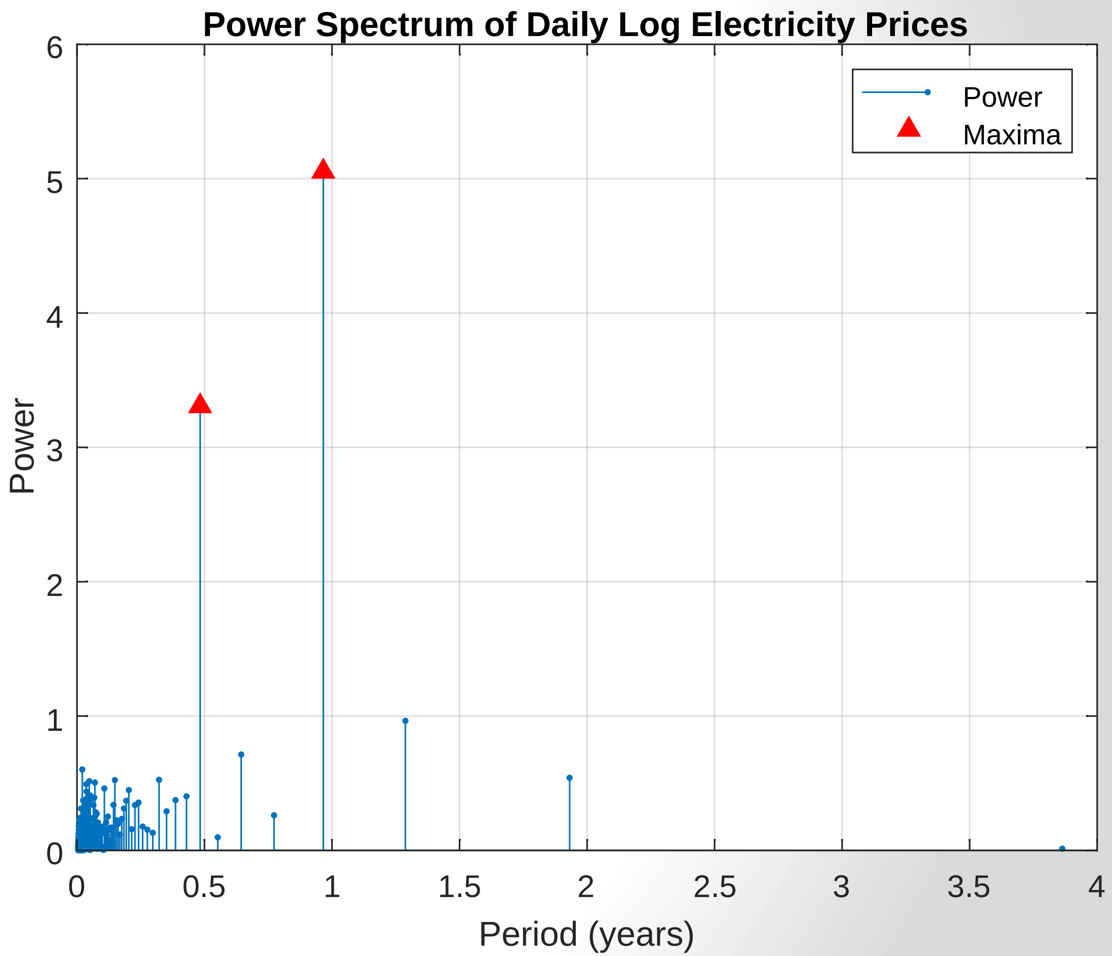
Long-Term Deterministic Trend



Identify Seasonal Components of the Long-Term Trend

1. Remove the linear trend from the log spot prices.
2. Apply the Fourier transform to the result.
3. Compute the power spectrum of the transformed data and the corresponding frequency vector.

Identify Seasonal Components of the Long-Term Trend



Identify Seasonal Components of the Long-Term Trend

- The dominant seasonal components correspond to 6-month and 12-month cycles.

Fit Long-Term Deterministic Trend

- By combining the dominant seasonal components, estimated by the spectral analysis, with the linear component, estimated by least squares, a linear model for the log spot prices can have this form:

$$\text{LogPrice}=\beta_0+\beta_1 t+\beta_2\sin(2\pi t)+\beta_3\cos(2\pi t)+\beta_4\sin(4\pi t)+\beta_5\cos(4\pi t)+\xi_t,$$

- t is years elapsed from the start of the sample.
- $\xi_t \sim N(0,\tau^2)$ is an iid sequence of random variables.
- β_2 and β_3 are the coefficients of the annual components.
- β_4 and β_5 are the coefficients of the semiannual components.

Fit Long-Term Deterministic Trend

- The sine and cosine terms in the model account for a possible phase offset. That is, for a phase offset θ:
- $A\sin(ft+\theta)=(A\cos\theta)\sin ft+(A\sin\theta)\cos ft=B\sin ft+C\cos ft$,
- where $B=A\cos\theta$ and $C=A\sin\theta$ are constants.
- Therefore, the inclusion of sine and cosine terms with the same frequency accounts for a phase offset.
- Stepwise regression is used to select the important predictors.

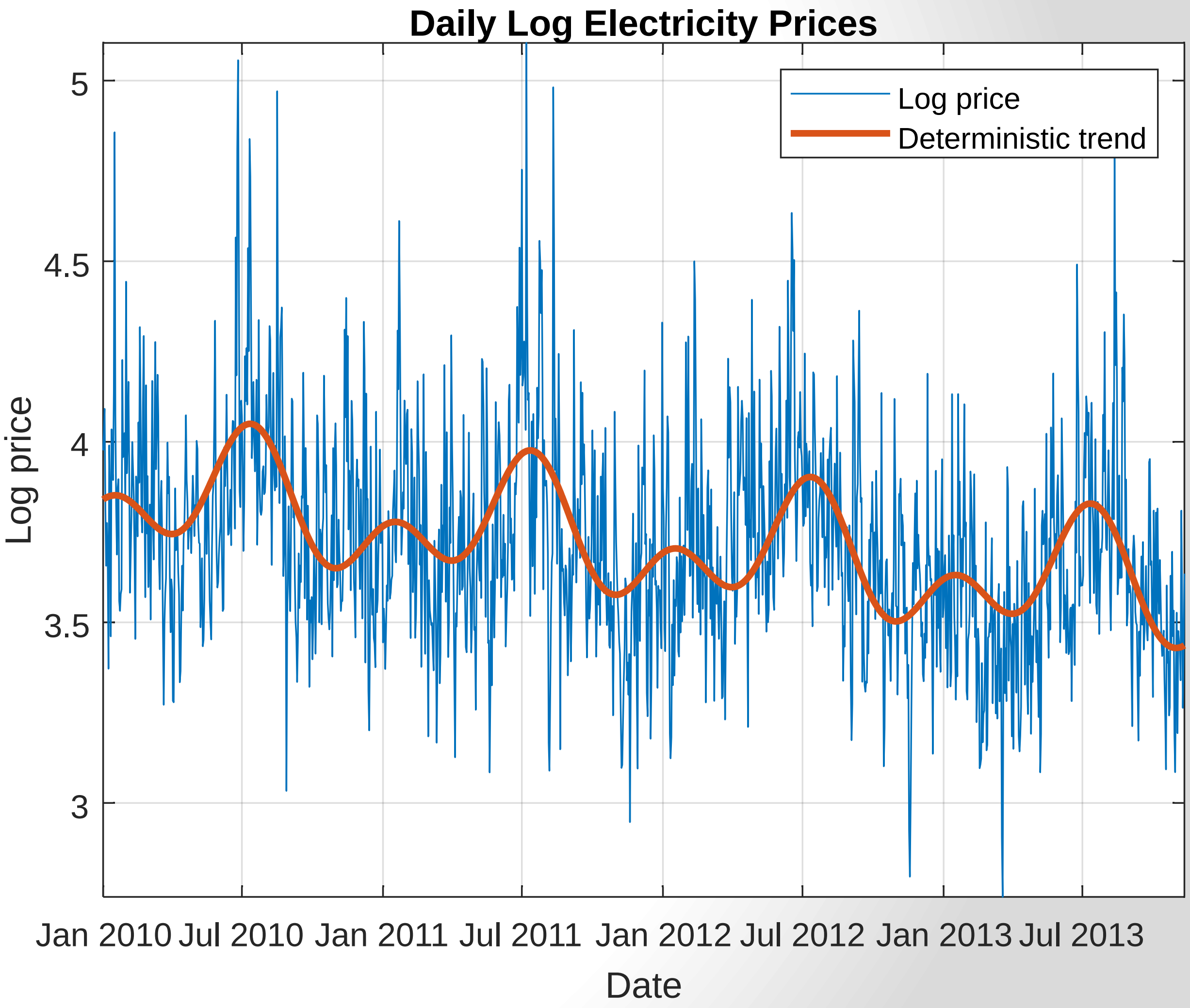
Fit Long-Term Deterministic Trend

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	3.8617	0.014114	273.6	0
t	-0.073594	0.0063478	-11.594	9.5312e-30
cos(2*pi*t)	-0.11982	0.010116	-11.845	6.4192e-31
sin(4*pi*t)	0.047563	0.0099977	4.7574	2.1629e-06
cos(4*pi*t)	0.098425	0.0099653	9.8768	2.7356e-22

Number of observations: 1411, Error degrees of freedom: 1406
Root Mean Squared Error: 0.264
R-squared: 0.221, Adjusted R-Squared: 0.219
F-statistic vs. constant model: 99.9, p-value = 7.15e-75

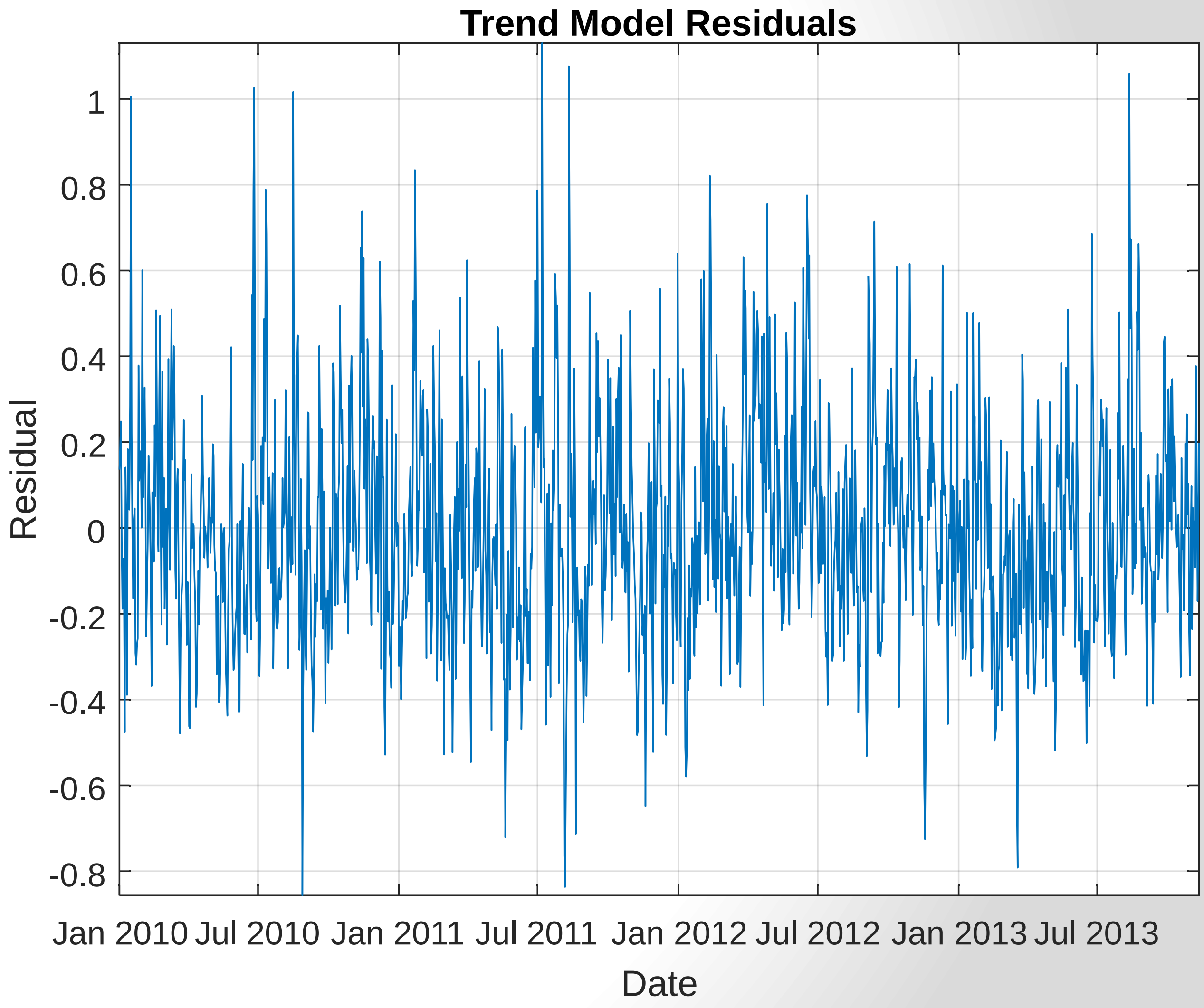
Fit Long-Term Deterministic Trend



Detrended Data

- A detrended time series is formed by removing the estimated long-term deterministic trend from the log spot prices. In other words, extract the residuals from the fitted stepwise linear regression model.

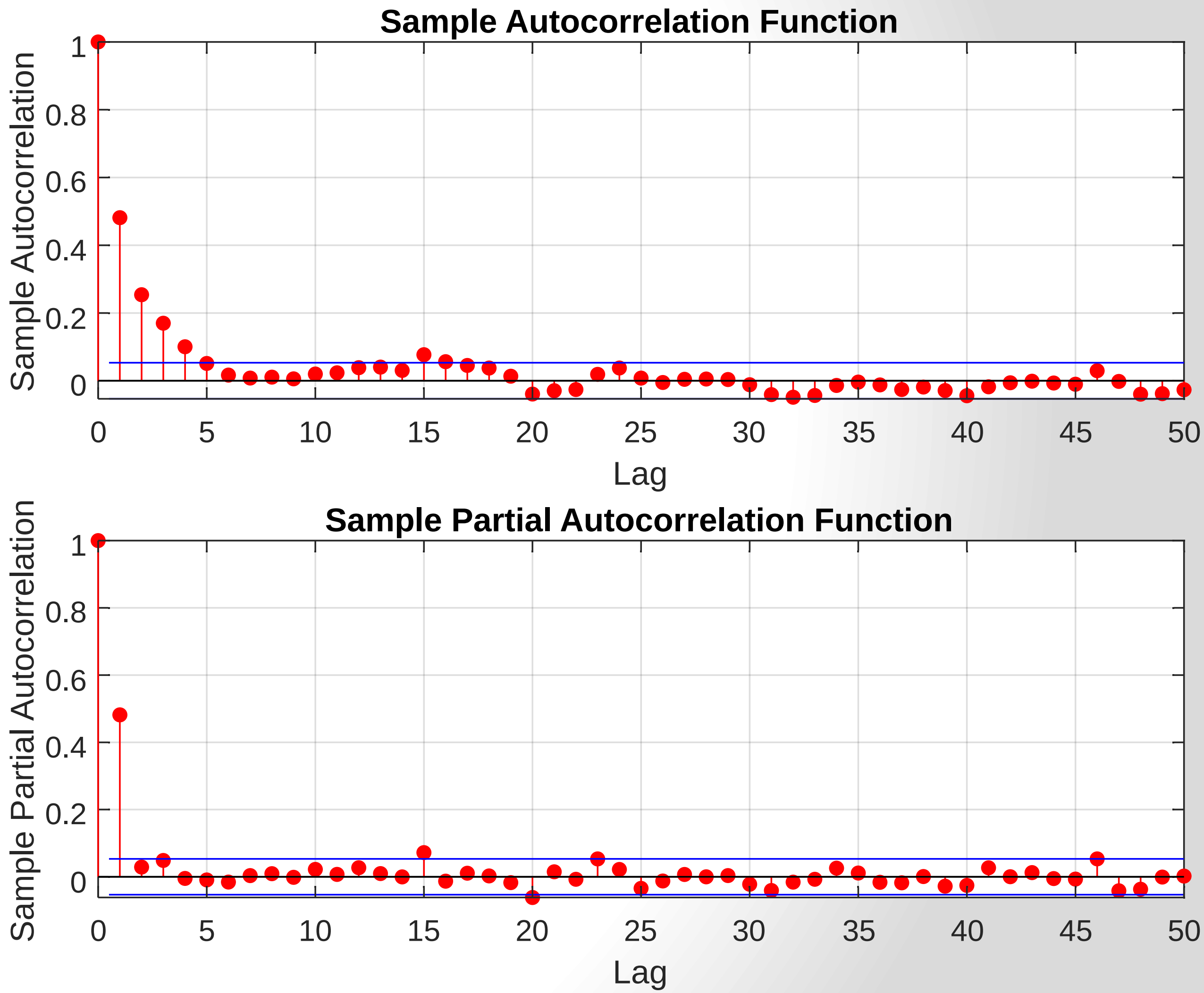
Detrended Data



Detrended Data

- The detrended data appears centered at zero, and the series exhibits serial autocorrelation because several clusters of consecutive residuals occur above and below y=0.
- These features suggest that an autoregressive model is appropriate for the detrended data.
- To determine the number of lags to include in the autoregressive model, apply the Box-Jenkins methodology. Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the detrended data in the same figure, but on separate plots.

Detrended Data



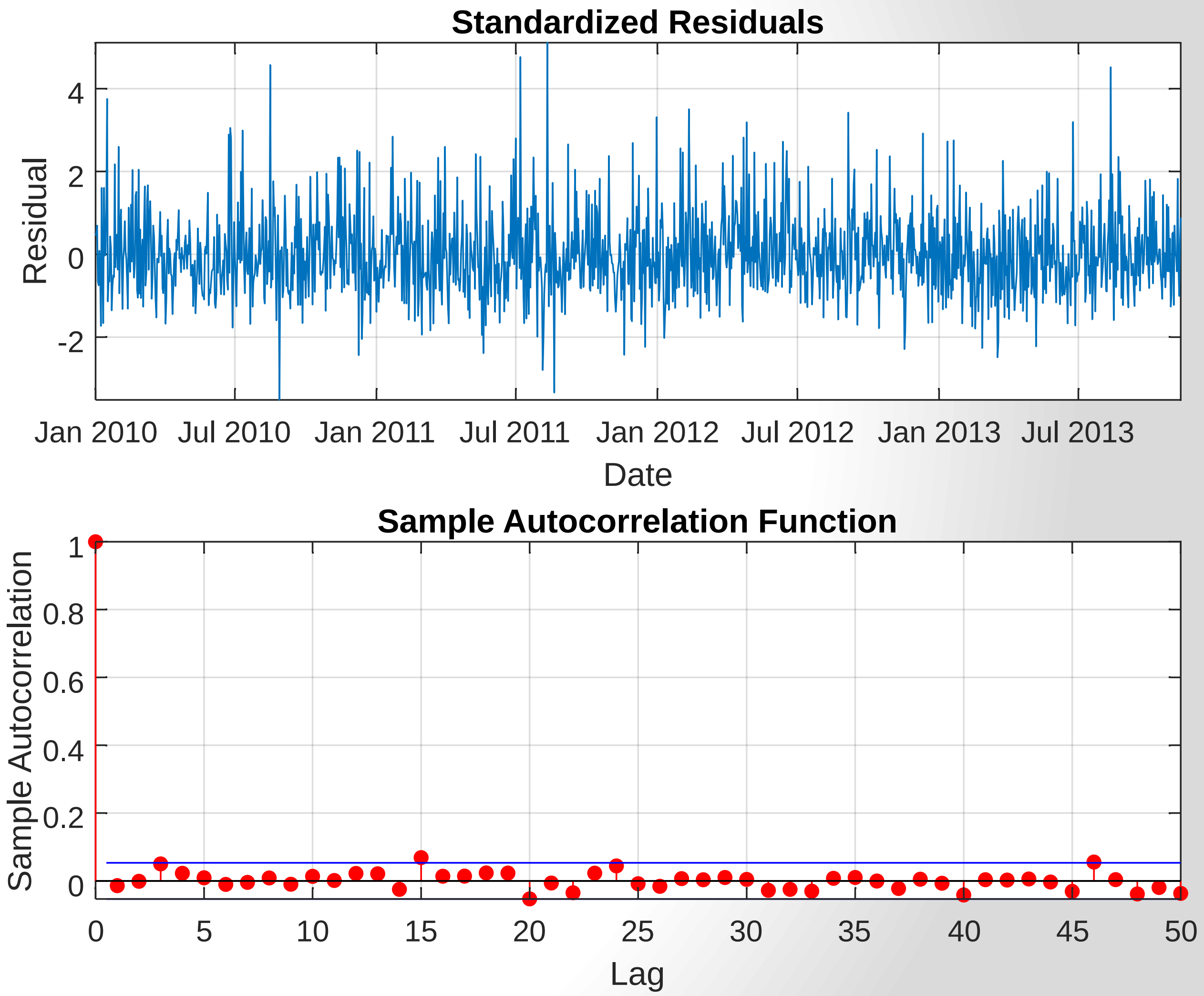
Detrended Data

The ACF gradually decays. The PACF exhibits a sharp cutoff after the first lag. This behavior is indicative of an AR(1) process.

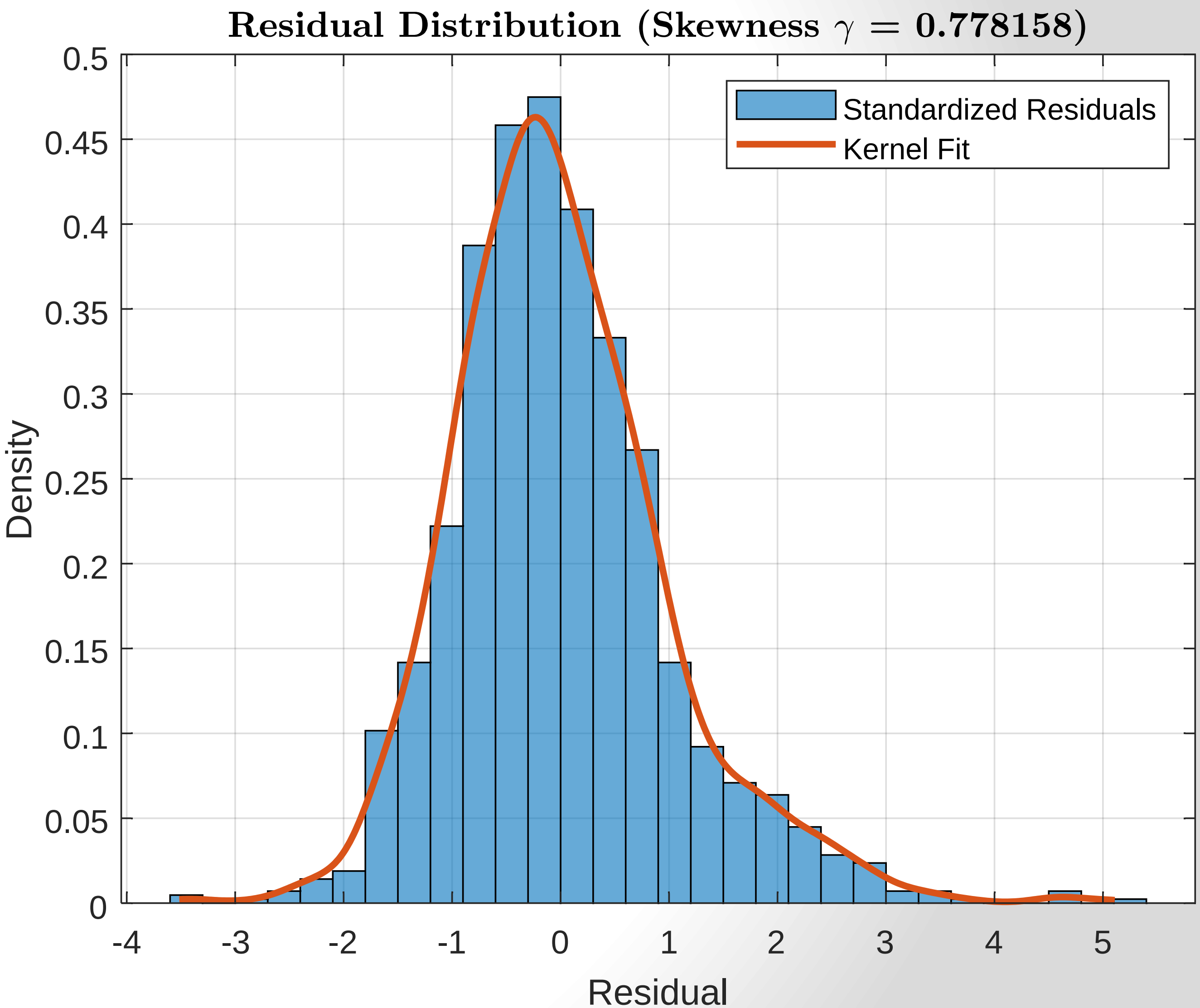
ARIMA(1,0,0) (Gaussian Distribution):

	Value	StandardError	TStatistic	PValue
Constant	0	0	NaN	NaN
AR{1}	0.4818	0.024353	19.784	4.0787e-87
Variance	0.053497	0.0014532	36.812	

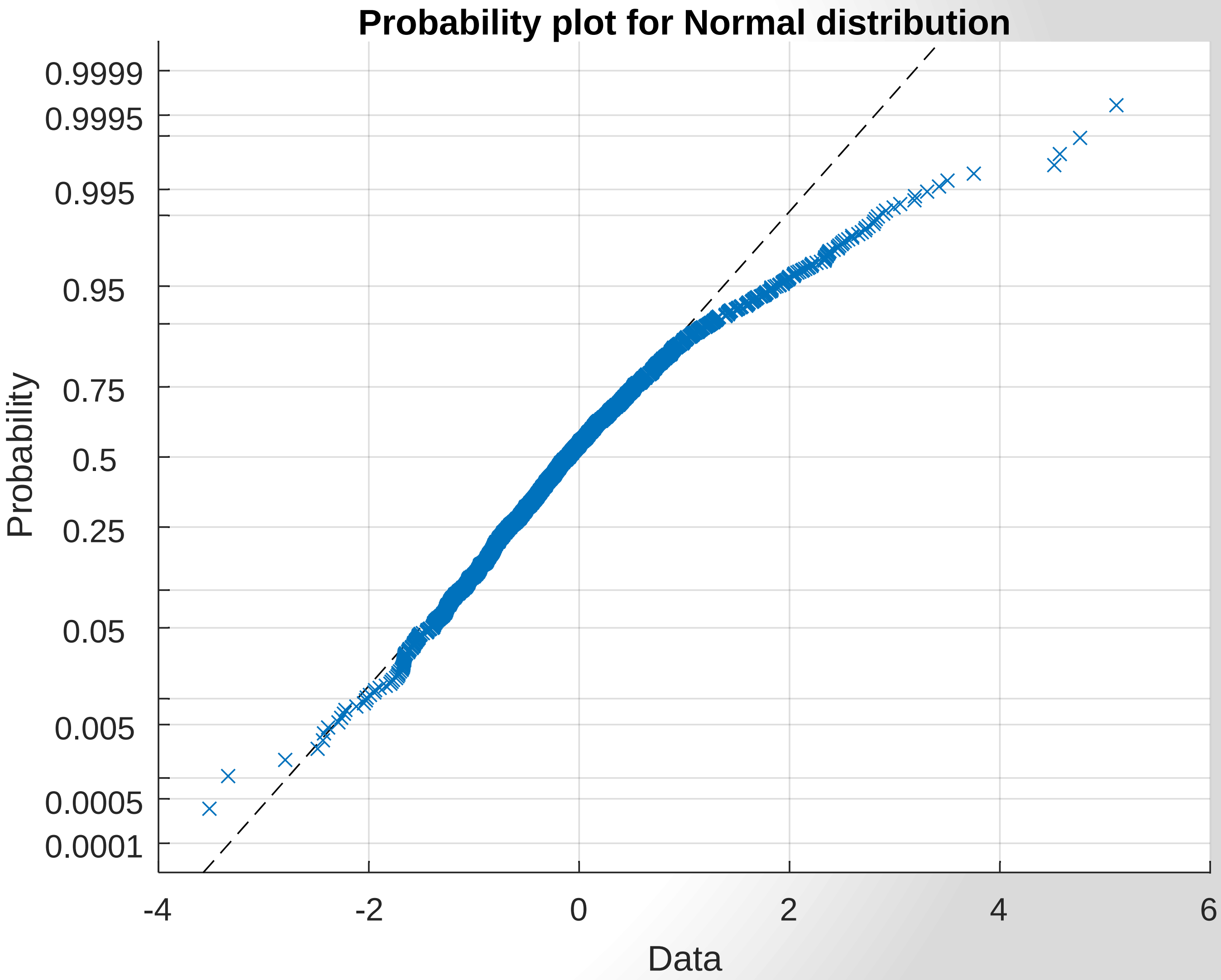
Detrended Data



Detrended Data



Detrended Data (Q-Q plot)



Detrended Data (Q-Q plot)

The residuals exhibit positive skewness because they deviate from normality in the upper tail.

Skewed Residual Series

- The epsilon-skew-normal distribution is a near-normal distribution family with location *theta*, scale σ, and additional skewness parameter ε .
- The skewness parameter models any nonzero skewness in the data.
- If ε=0, the epsilon-skew-normal distribution reduces to the normal distribution.

Skewed Residual Series

EpsilonSkewNormal distribution

theta = -0.421946

sigma = 0.972487

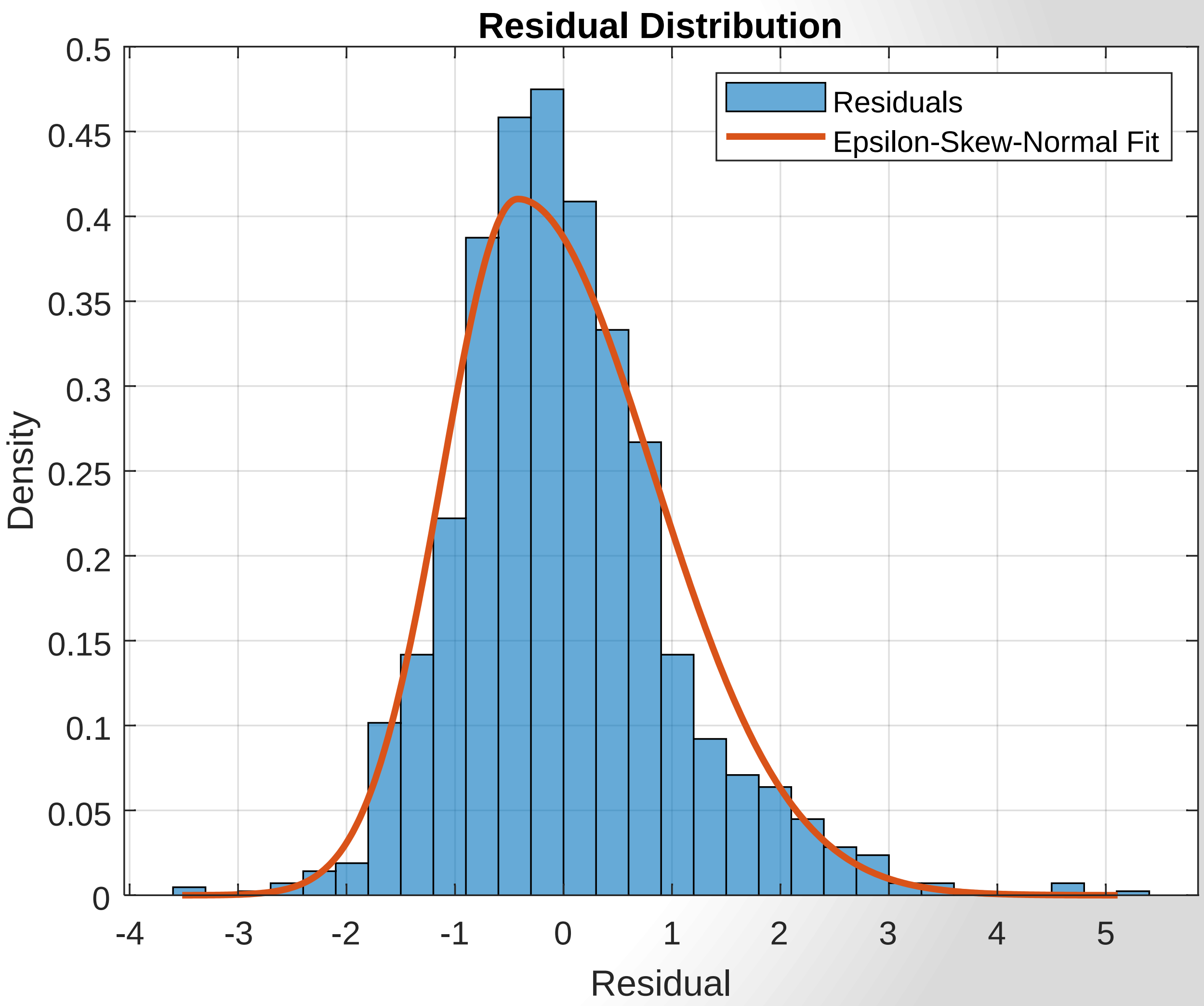
epsilon = -0.286248

Skewed Residual Series

Estimated parameter standard errors:

Parameter	StandardError
{'theta' }	0.0638
{'sigma' }	0.035606
{'epsilon'}	0.037877

Skewed Residual Series

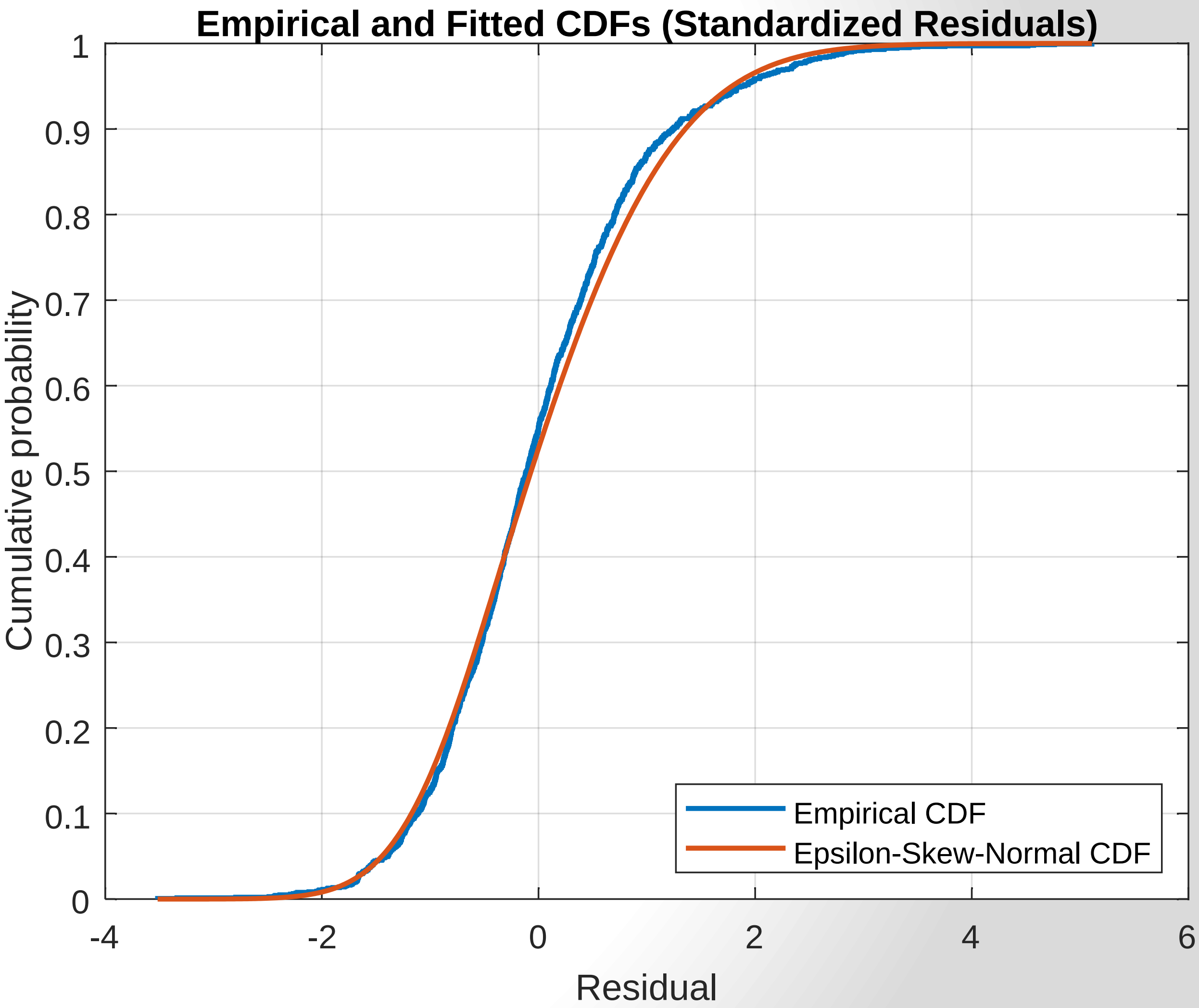


Goodness of Fit

the goodness of fit of the epsilon-skew-normal distribution :

1. Compare the empirical and fitted (reference) cumulative distribution functions (cdfs).
2. Conduct the Kolmogorov-Smirnov test for goodness of fit.
3. Plot the maximum cdf discrepancy.

Goodness of Fit

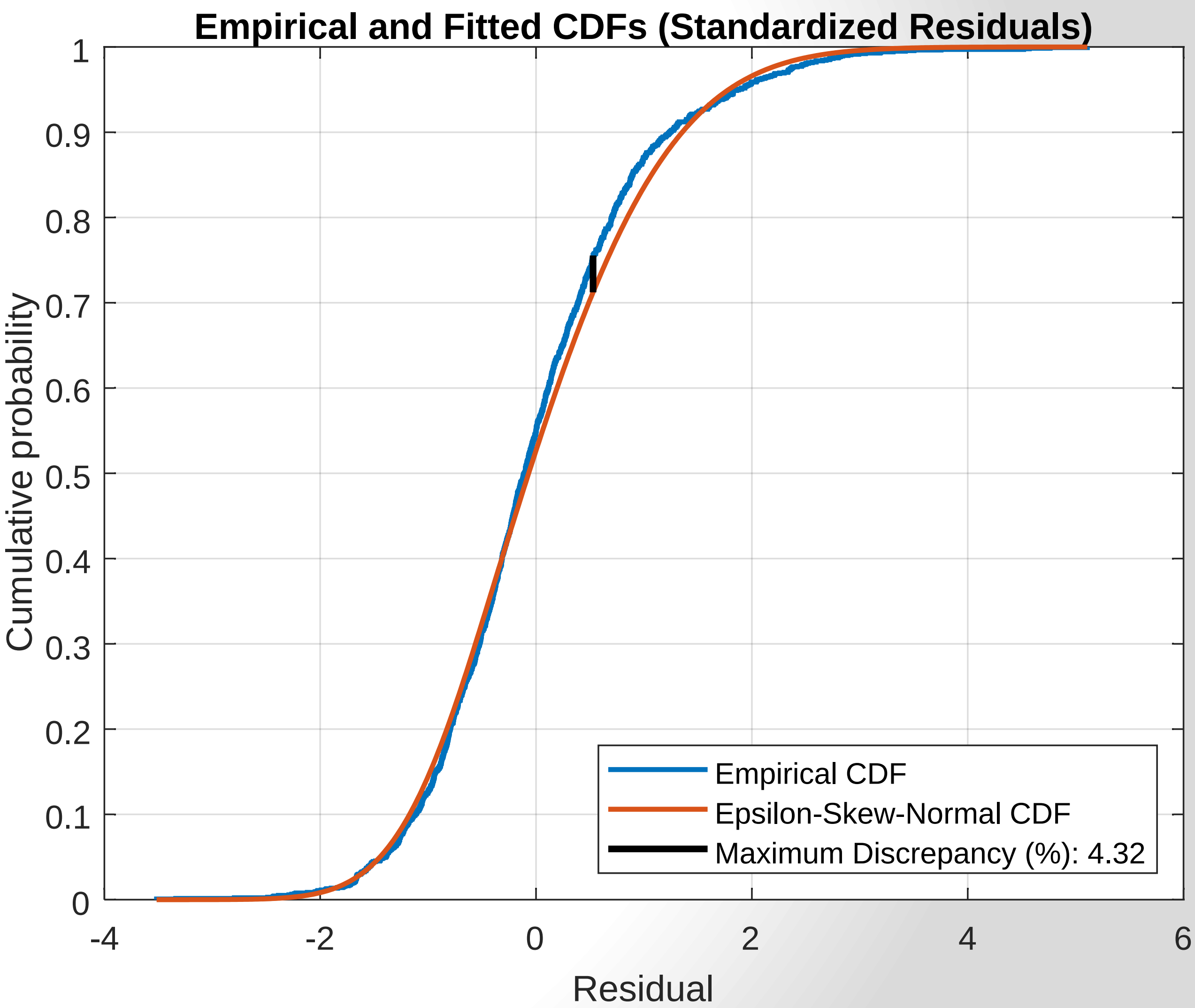


Goodness of Fit

Kolmogorov-Smirnov test p-value is 0.85439.

The p-value of the test is large enough to suggest that the null hypothesis that the distributions are the same should not be rejected.

Goodness of Fit



Simulate Future Electricity Spot Prices

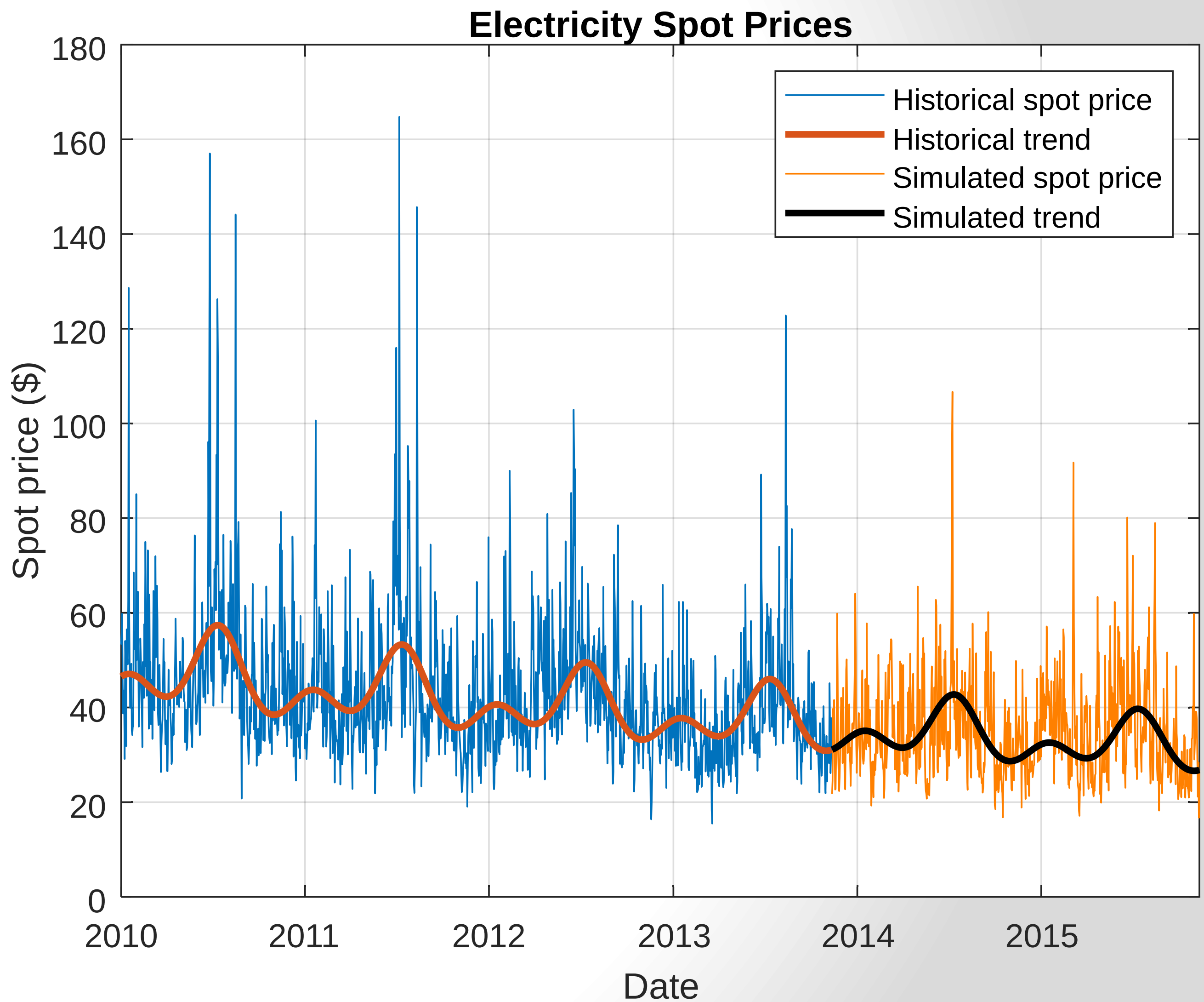
Simulate future electricity spot prices over the two-year horizon following the end of the historical data. Construct a model from which to simulate, composed of the estimated components of the time series:

1. Specify the dates for the forecast horizon.
2. Obtain simulated residuals by simulating standardized residuals from the fitted epsilon-skew-normal distribution, then scaling the result by the estimated instantaneous standard deviations.
3. Obtain simulated, detrended log prices by filtering the residuals through the fitted AR(1) model.

Simulate Future Electricity Spot Prices

4. Forecast values of the deterministic trend using the fitted model.
5. Obtain simulated log prices by combining the simulated, detrended log prices and the forecasted, deterministic trend values.
6. Obtain simulated spot prices by exponentiating the simulated log spot prices.

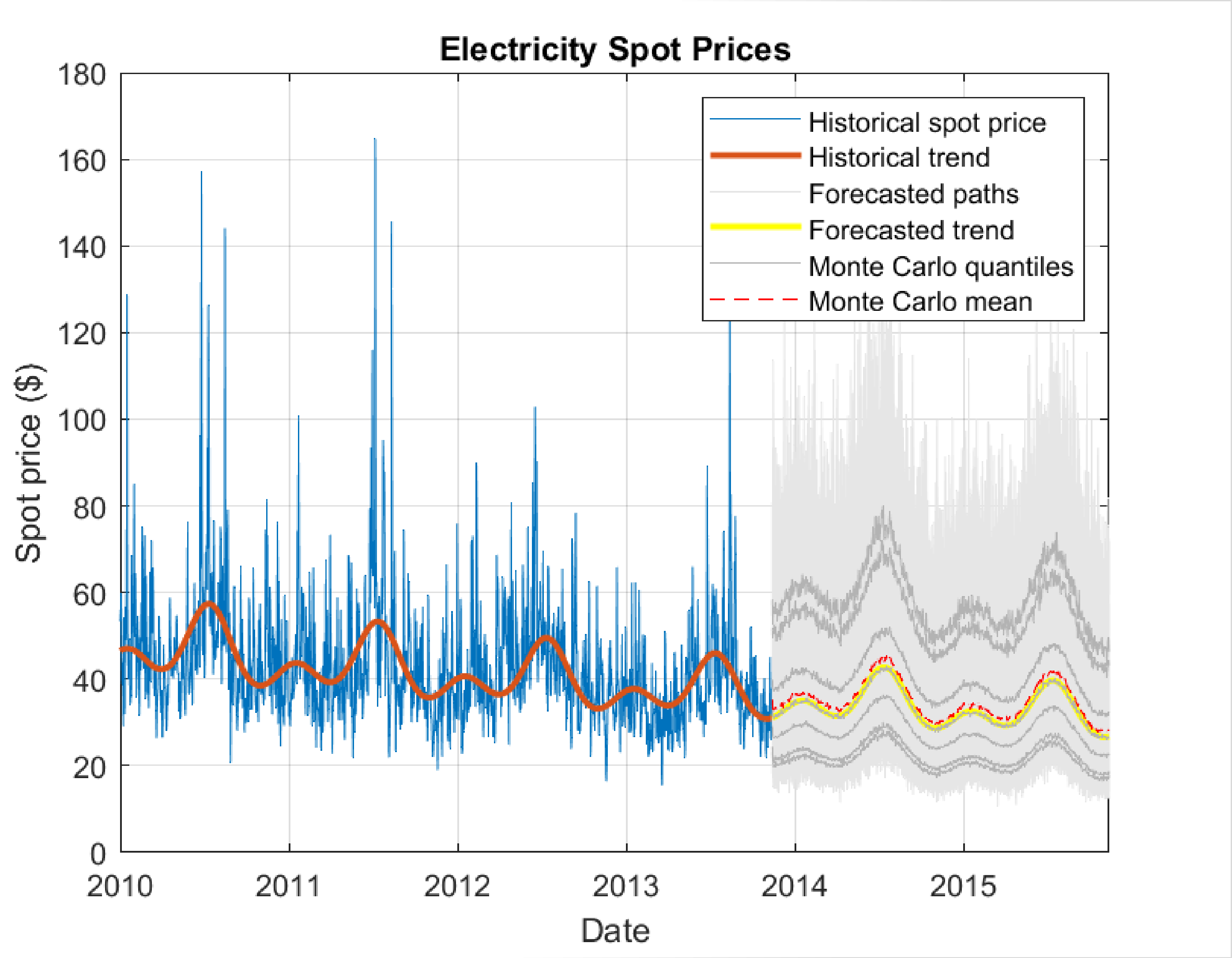
Simulated Paths



Simulated Paths

The Monte Carlo statistics is obtained from the simulated spot price paths by computing, for each time point in the forecast horizon, the mean and median, and the 2.5th, 5th, 25th, 75th, 95th, and 97.5th percentiles.

Simulated Paths

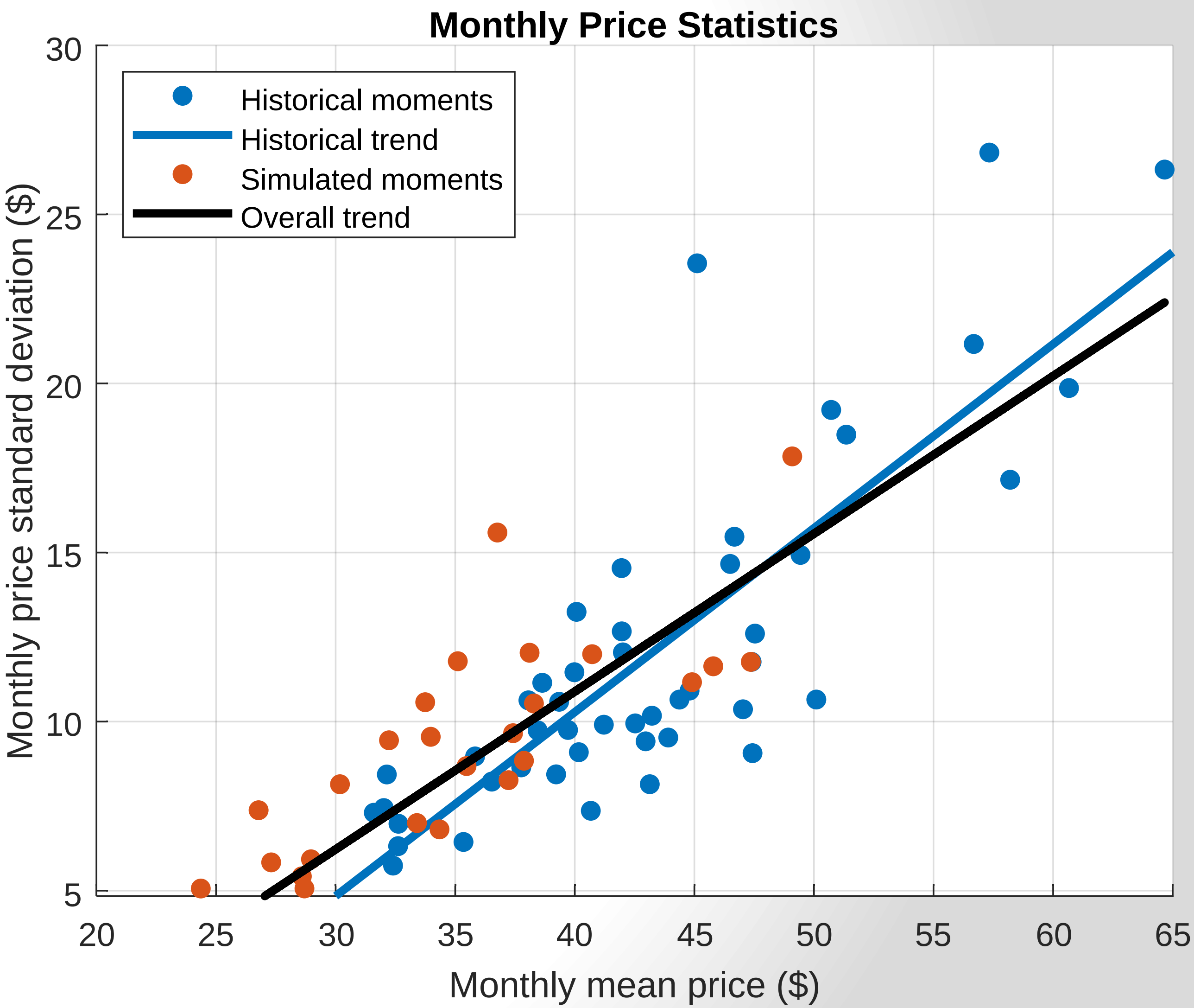


Simulated Paths

Assess whether the model addresses the large spikes exhibited in the historical data:

1. Estimate the monthly Monte Carlo moments of the simulated spot price paths.
2. Plot a line of best fit to the historical monthly moments.
3. Plot a line of best fit to the combined historical and Monte Carlo moments.
4. Visually compare the lines of best fit.

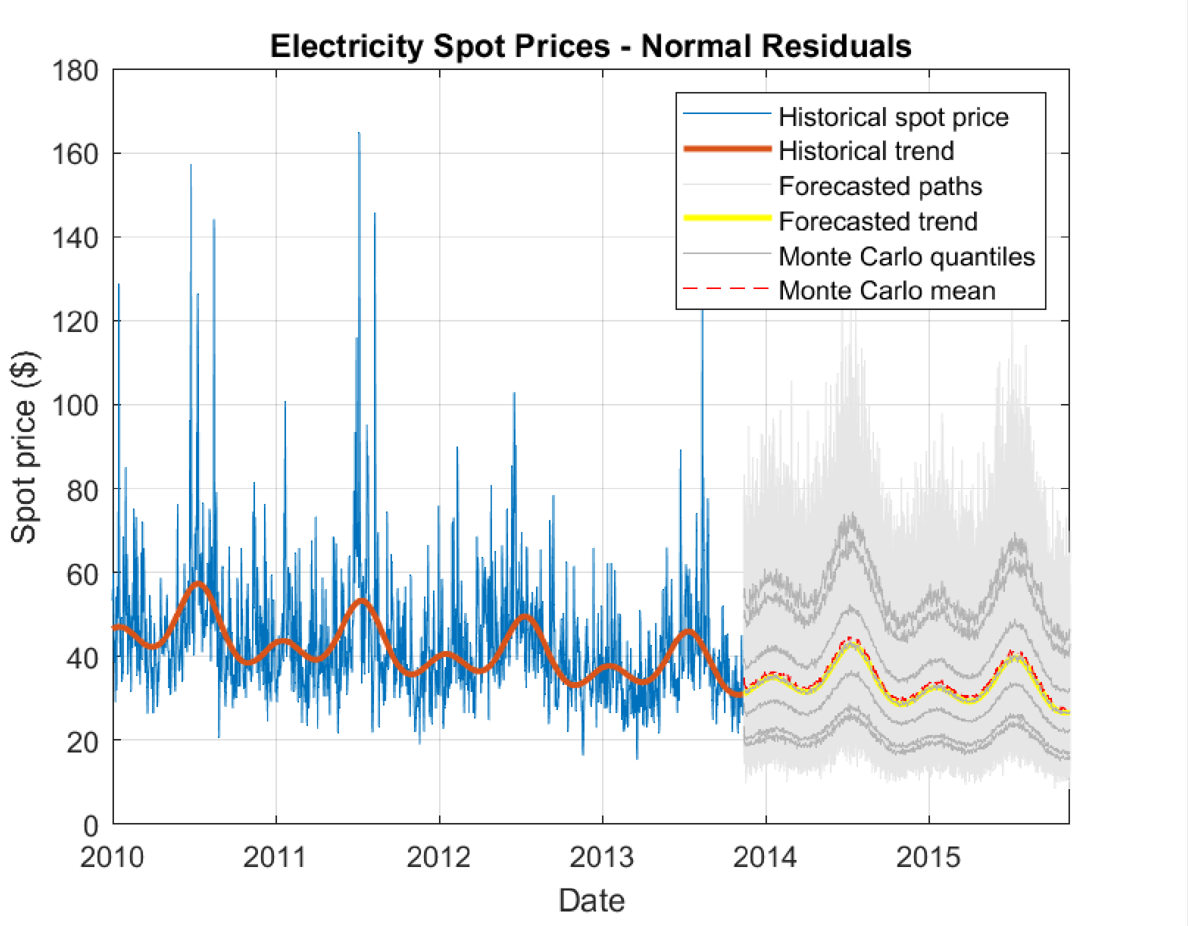
Simulated Paths



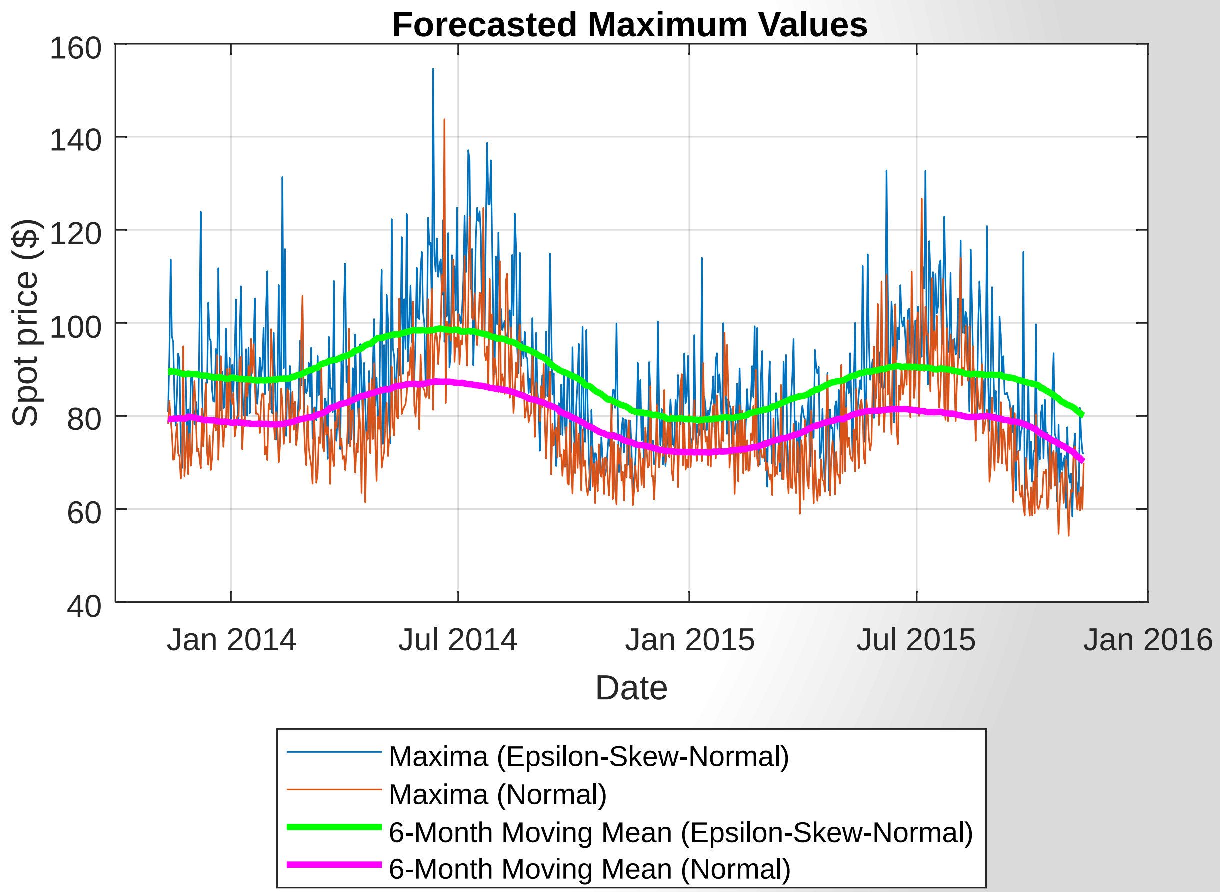
Simulated Paths

- The simulated monthly moments are broadly consistent with the historical data.
- The simulated spot prices tend to exhibit smaller spikes than the observed spot prices.
- To account for larger spikes, you can model a larger right tail by applying extreme value theory in the distribution fitting step.
- This approach uses an epsilon-skew-normal distribution for the residuals, but models the upper tail by using the generalized Pareto distribution.

Compare Epsilon-Skew-Normal Results with Normal Distribution Assumption



Compare Epsilon-Skew-Normal Results with Normal Distribution Assumption



Compare Epsilon-Skew-Normal Results with Normal Distribution Assumption

Although the series of simulated maxima are correlated, the plot shows that the epsilon-skew-normal maxima are larger than the normal maxima.