In [59]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import lag_plot
```

In [60]:
```
df = pd.read_csv("./14100355.csv", usecols = ['REF_DATE', 'GEO', 'North American Indust
ds =  pd.read_csv("./14100355.csv", usecols = ['REF_DATE', 'GEO', 'North American Indus
```

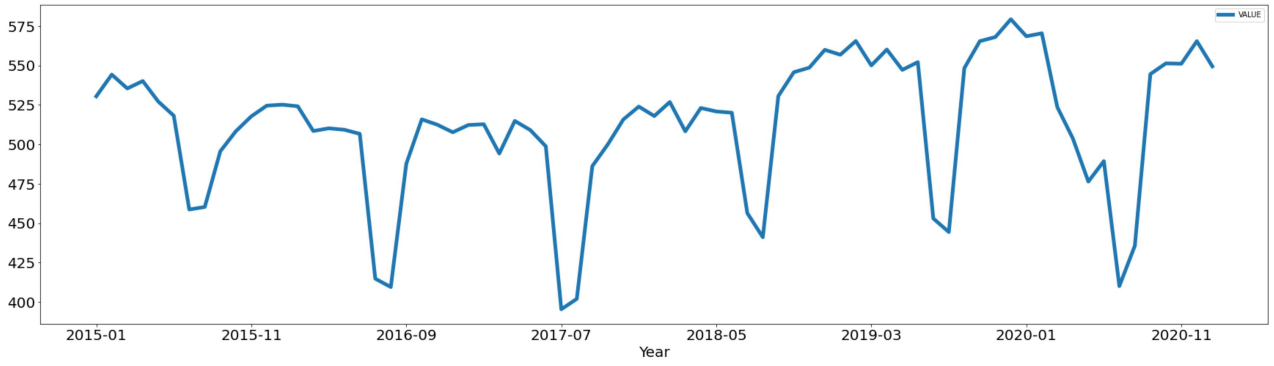
In [61]:
```
df = df.rename(columns = {"North American Industry Classification System (NAICS)":"Indu
ds = ds.rename(columns = {"North American Industry Classification System (NAICS)":"Indu
```

In [62]:
```
df = df[(df['GEO'] =='Ontario') & (df['Industry'] == 'Educational services [61]') & (df
ds = ds[(ds['GEO'] =='Ontario') & (ds['Industry'] == 'Health care and social assistance
```

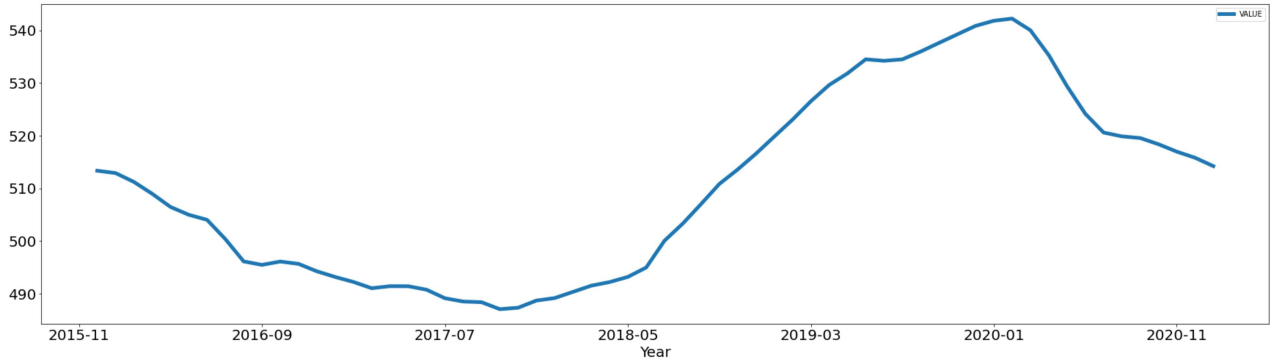
In [63]:
```
df.month = pd.to_datetime(df.REF_DATE)
df.set_index('REF_DATE', inplace=True)

ds.month = pd.to_datetime(ds.REF_DATE)
ds.set_index('REF_DATE', inplace=True)
```

```
<ipython-input-63-68f3463cb3fb>:1: UserWarning: Pandas doesn't allow columns to be creat
ed via a new attribute name - see https://pandas.pydata.org/pandas-docs/stable/indexing.
html#attribute-access
  df.month = pd.to_datetime(df.REF_DATE)
<ipython-input-63-68f3463cb3fb>:4: UserWarning: Pandas doesn't allow columns to be creat
ed via a new attribute name - see https://pandas.pydata.org/pandas-docs/stable/indexing.
html#attribute-access
  ds.month = pd.to_datetime(ds.REF_DATE)
```

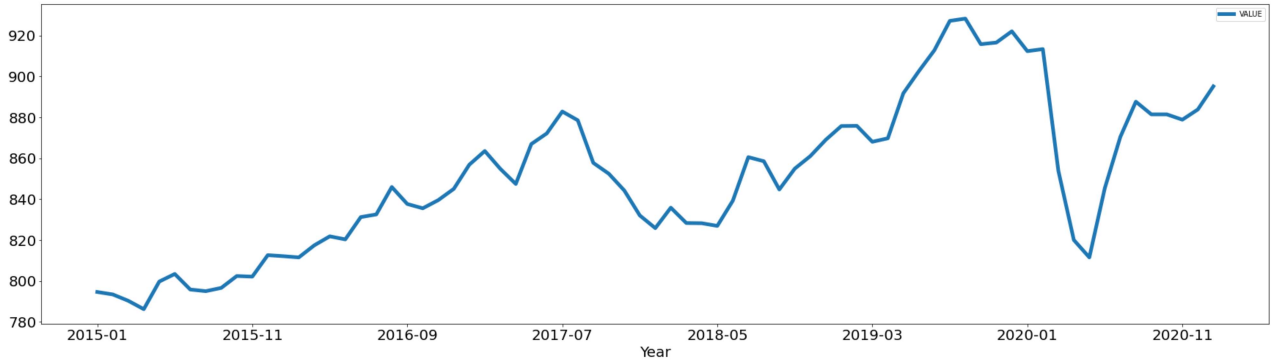
In [64]:
```
df.plot(figsize=(30,8), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



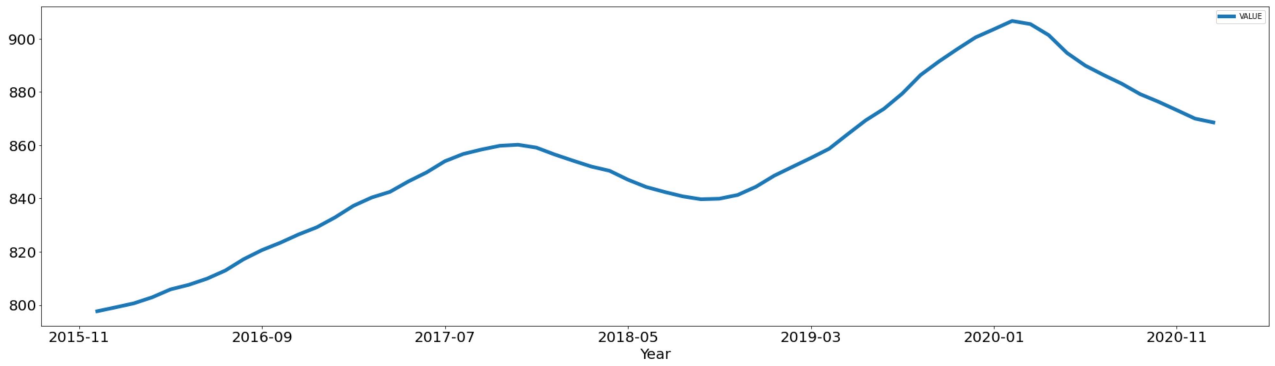
In [65]:
```
val1 = df[['VALUE']]
val1.rolling(12).mean().plot(figsize=(30,8), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



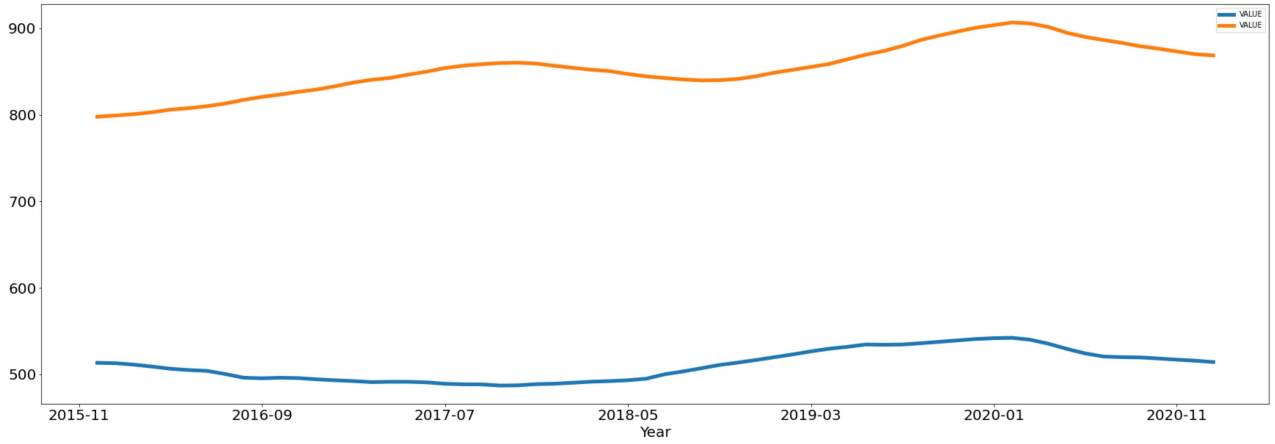
In [66]:
```
ds.plot(figsize=(30,8), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



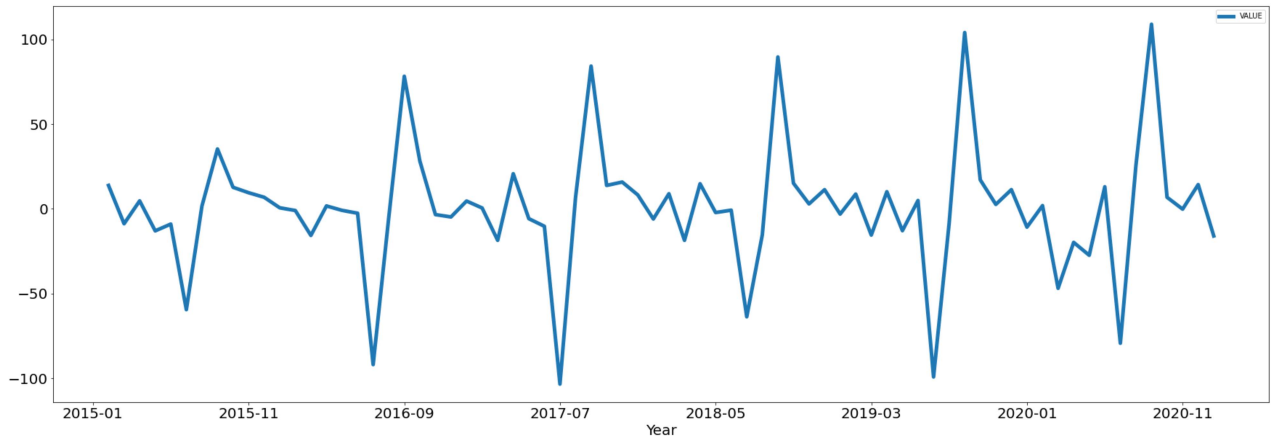
In [67]:
```
val2 = ds[['VALUE']]
val2.rolling(12).mean().plot(figsize=(30,8), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



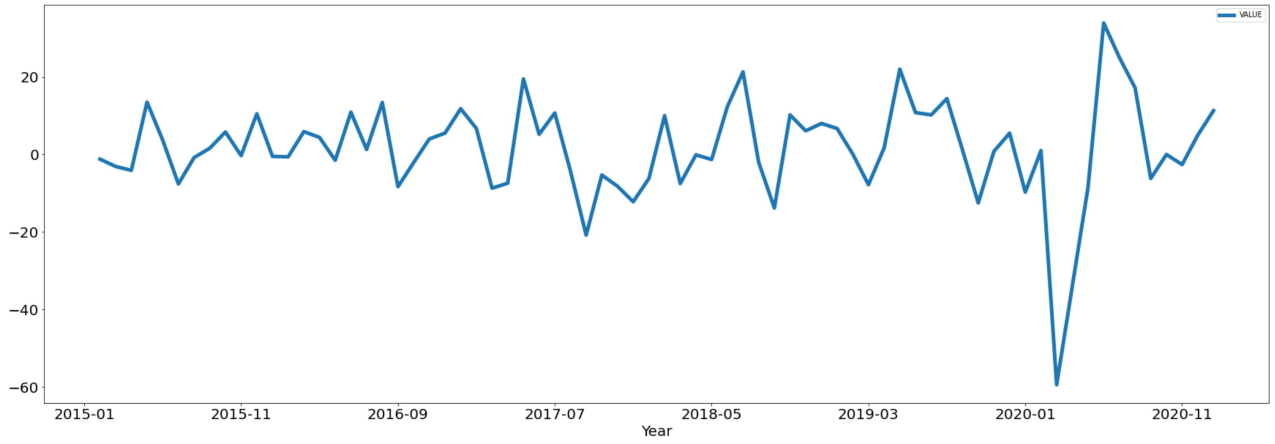
In [68]:
```
df_rm = pd.concat([val1.rolling(12).mean(), val2.rolling(12).mean()], axis=1)
df_rm.plot(figsize=(30,10), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



In [69]:
```
val1.diff().plot(figsize=(30,10), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



In [70]:
```
val2.diff().plot(figsize=(30,10), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```



In [78]:
```
sns.lmplot(x='REF_DATE', y='VALUE', fit_reg=False, data=ds')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-78-7efdce4ee938> in <module>
----> 1 sns.lmplot(x='REF_DATE', y='VALUE', fit_reg=False, data=ds, hue='VALUE')

~\anaconda3\lib\site-packages\seaborn\_decorators.py in inner_f(*args, **kwargs)
     44             )
     45         kwargs.update({k: arg for k, arg in zip(sig.parameters, args)})
---> 46         return f(**kwargs)
```

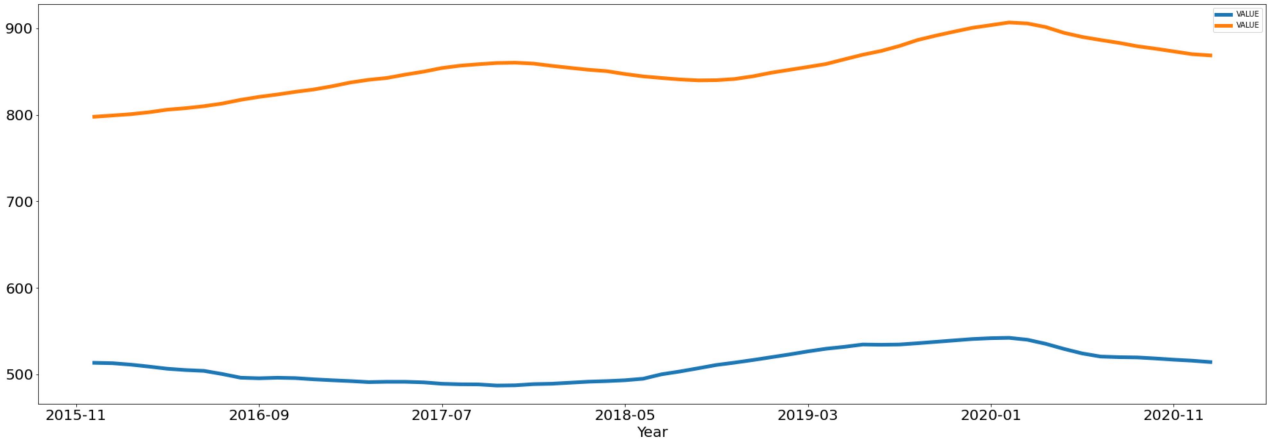
```
     47     return inner_f
     48 

~\anaconda3\lib\site-packages\seaborn\regression.py in lmplot(x, y, data, hue, col, row,
palette, col_wrap, height, aspect, markers, sharex, sharey, hue_order, col_order, row_or
der, legend, legend_out, x_estimator, x_bins, x_ci, scatter, fit_reg, ci, n_boot, units,
seed, order, logistic, lowess, robust, logx, x_partial, y_partial, truncate, x_jitter, y
_jitter, scatter_kws, line_kws, size)
    583     need_cols = [x, y, hue, col, row, units, x_partial, y_partial]
    584     cols = np.unique([a for a in need_cols if a is not None]).tolist()
--> 585     data = data[cols]
    586 
    587     # Initialize the grid

~\anaconda3\lib\site-packages\pandas\core\frame.py in __getitem__(self, key)
   2906             if is_iterator(key):
   2907                 key = list(key)
-> 2908             indexer = self.loc._get_listlike_indexer(key, axis=1, raise_missing
=True)[1]
   2909 
   2910         # take() does not accept boolean indexers

~\anaconda3\lib\site-packages\pandas\core\indexing.py in _get_listlike_indexer(self, ke
y, axis, raise_missing)
   1252             keyarr, indexer, new_indexer = ax._reindex_non_unique(keyarr)
   1253 
-> 1254         self._validate_read_indexer(keyarr, indexer, axis, raise_missing=raise_
missing)
   1255         return keyarr, indexer
   1256 

~\anaconda3\lib\site-packages\pandas\core\indexing.py in _validate_read_indexer(self, ke
y, indexer, axis, raise_missing)
   1302             if raise_missing:
   1303                 not_found = list(set(key) - set(ax))
-> 1304                 raise KeyError(f"{not_found} not in index")
   1305 
   1306             # we skip the warning on Categorical

KeyError: "['REF_DATE'] not in index"
```

In [82]:
```
df_rm.corr()
```

Out[82]:

	VALUE	VALUE
VALUE	1.000000	0.603728
VALUE	0.603728	1.000000

In [84]:
```
df_rm.plot(figsize=(30,10), linewidth=5, fontsize=20)
plt.xlabel('Year', fontsize=20);
```

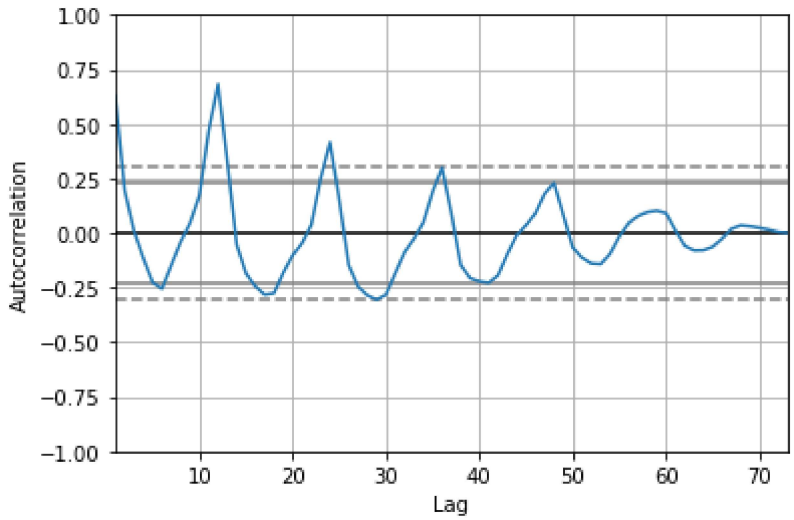


In [85]:
```
df_rm.diff().corr()
```

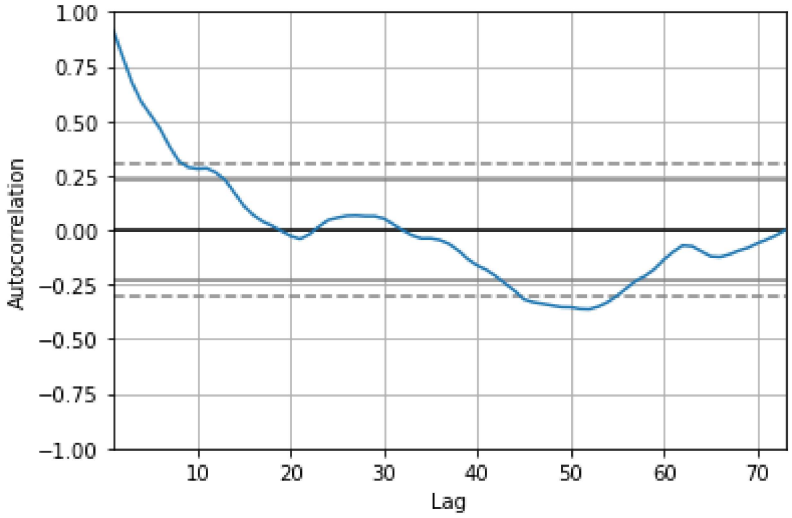
Out[85]:

	VALUE	VALUE
VALUE	1.000000	0.276499
VALUE	0.276499	1.000000

In [86]:
```
pd.plotting.autocorrelation_plot(val1);
```



In [87]:
```
pd.plotting.autocorrelation_plot(val2);
```


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Autocorrelation
10
20
30
40
50
60
70
Lag

In [ ]: