Bitcoin Linear Regression

Introduction

- The USD for example is backed by trust in the US Government and as the world's reserve currency.
  - US Government continues to put more currency into circulation.
  - More currency into circulation devalues the currency over time.
  - USD is not currently backed by anything of physical value other than the US government
- The value of BTC is merely based on the law of supply and demand.
  - BTC has a cap on the amount of coins that can be produced, 21 million.
  - Due to the supply being limited to 21 million, as demand increases the value will likely increase.

- The likelihood of the United States adopting BTC as its universal currency is small
- Some countries that are susceptible to significant currency manipulation, like hyperinflation, may be enticed to use bitcoin.
- Individuals may revert to BTC so that their money holds value

Goal

Identify features around the Bitcoin blockchain that might have an impact on the price of bitcoin, such as, market capitalization, transaction volume, miners revenue, transactions per block, estimated volume, average block size, hash rate, number of orphan blocks...

Proposed Method

A multiple linear regression model where multiple explanatory variables will help us in predicting the price of bitcoin. Several independent variables are explored and analyzed. After the analysis, the most correlated variables with the price of bitcoin are selected. First, the dataset is read and its data is pre-processed. Next, we examine the historical value of bitcoin over time. After that, we will review the correlation between variables to see which one will be included in the linear regression analysis.

In [1]:

```
!pip install requests pandas numpy matplotlib seaborn sklearn statsmodels
```

```
Requirement already satisfied: requests in /Users/alvaroserranorivas/.pyenv/vers
ions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (2.26.0)
Requirement already satisfied: pandas in /Users/alvaroserranorivas/.pyenv/versio
ns/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (1.3.4)
Requirement already satisfied: numpy in /Users/alvaroserranorivas/.pyenv/version
```

```
s/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (1.21.4)
Requirement already satisfied: matplotlib in /Users/alvaroserranorivas/.pyenv/ve
rsions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (3.4.3)
Requirement already satisfied: seaborn in /Users/alvaroserranorivas/.pyenv/versi
ons/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (0.11.2)
Requirement already satisfied: sklearn in /Users/alvaroserranorivas/.pyenv/versi
ons/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (0.0)
Requirement already satisfied: statsmodels in /Users/alvaroserranorivas/.pyenv/v
ersions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (0.13.
0)
Requirement already satisfied: charset-normalizer~=2.0.0; python_version ⩾ "3"
in /Users/alvaroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regressio
n/lib/python3.9/site-packages (from requests) (2.0.7)
Requirement already satisfied: certifi⩾2017.4.17 in /Users/alvaroserranorivas/.
pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages
(from requests) (2021.10.8)
Requirement already satisfied: idna<4,⩾2.5; python_version ⩾ "3" in /Users/alv
aroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python
3.9/site-packages (from requests) (3.3)
Requirement already satisfied: urllib3<1.27,⩾1.21.1 in /Users/alvaroserranoriva
s/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packag
es (from requests) (1.26.7)
Requirement already satisfied: python-dateutil⩾2.7.3 in /Users/alvaroserranoriv
as/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packa
ges (from pandas) (2.8.2)
Requirement already satisfied: pytz⩾2017.3 in /Users/alvaroserranorivas/.pyenv/
versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from
pandas) (2021.3)
Requirement already satisfied: pyparsing⩾2.2.1 in /Users/alvaroserranorivas/.py
env/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (f
rom matplotlib) (3.0.4)
Requirement already satisfied: kiwisolver⩾1.0.1 in /Users/alvaroserranorivas/.p
yenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages
(from matplotlib) (1.3.2)
Requirement already satisfied: cycler⩾0.10 in /Users/alvaroserranorivas/.pyenv/
versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from
matplotlib) (0.11.0)
Requirement already satisfied: pillow⩾6.2.0 in /Users/alvaroserranorivas/.pyen
v/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (fro
m matplotlib) (8.4.0)
Requirement already satisfied: scipy⩾1.0 in /Users/alvaroserranorivas/.pyenv/ve
rsions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from se
aborn) (1.7.2)
Requirement already satisfied: scikit-learn in /Users/alvaroserranorivas/.pyenv/
versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from
sklearn) (1.0.1)
Requirement already satisfied: patsy⩾0.5.2 in /Users/alvaroserranorivas/.pyenv/
versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from
statsmodels) (0.5.2)
Requirement already satisfied: six⩾1.5 in /Users/alvaroserranorivas/.pyenv/vers
ions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from pyth
on-dateutil⩾2.7.3→pandas) (1.16.0)
Requirement already satisfied: threadpoolctl⩾2.0.0 in /Users/alvaroserranoriva
s/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packag
```

```
es (from scikit-learn->sklearn) (3.0.0)
Requirement already satisfied: joblib>=0.11 in /Users/alvaroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/lib/python3.9/site-packages (from scikit-learn->sklearn) (1.1.0)
WARNING: You are using pip version 20.2.3; however, version 21.3.1 is available.
You should consider upgrading via the '/Users/alvaroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/bin/python3.9 -m pip install --upgrade pip' command.
```

In [2]:

```
import requests
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
from datetime import datetime
import time
import statsmodels.formula.api as smf
import scipy.stats as stats
import statsmodels.api as sm
import matplotlib.cm as cm
from statsmodels.graphics.gofplots import ProbPlot
from sklearn import linear_model
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
sns.set(style="whitegrid")
```

In [3]:

```
USD= 2208000000000 # USD in circulation as of Nov 6, 2021 according to https://ycharts.com/indicators/us_currency_in_circulation
Bitcoin=21000000   # 21 Million Bitcoin (Most Bitcoin that will be in Circulation)
Value=USD/Bitcoin  # Value of Bitcoin compared to the current dollar
print(f"USD/BTC in circulation: {Value}")
# Get current USD/BTC price from https://www.coindesk.com/price/
session = requests.Session()
```

```
USD_BTC_rate =
session.get("https://api.coindesk.com/v1/bpi/currentprice.json").
["bpi"]["USD"]["rate_float"]
print(f"Current USD/BTC rate: {USD_BTC_rate}")
```

```
USD/BTC in circulation: 105142.85714285714
Current USD/BTC rate: 64648.3643
```

Data set description

In [4]:

```
BTC_data = pd.read_csv("bitcoin_dataset.csv", header=0)
print(BTC_data.head())
print(BTC_data.columns)
print(BTC_data.shape)
BTC_data["Date"] = BTC_data["Date"].apply(lambda x:
datetime.strptime(x, '%Y-%m-%d %H:%M:%S'))
# Create a Days column that is the number of days for each row
BTC_data["Days"] = (BTC_data["Date"] -
BTC_data["Date"].min()).dt.days
# Print subset of data where Median Confirmation Time is greater
than 0
print(BTC_data[BTC_data["btc_median_confirmation_time"] > 0])
# Subset where Median Confirmation Time is greater than 0
BTC_data2 = BTC_data[BTC_data["btc_median_confirmation_time"] >
0]
# Drop NA values
BTC_data2.dropna(inplace=True)
```

```
                 Date  btc_market_price  btc_total_bitcoins  btc_market_cap  \
0  2009-11-10 00:00:00               0.0           1339450.0             0.0
1  2009-11-11 00:00:00               0.0           1342900.0             0.0
2  2009-11-12 00:00:00               0.0           1346400.0             0.0
3  2009-11-13 00:00:00               0.0           1349900.0             0.0
4  2009-11-14 00:00:00               0.0           1354050.0             0.0

   btc_trade_volume  btc_blocks_size  btc_avg_block_size  \
0               0.0              0.0            0.000215
1               0.0              0.0            0.000323
2               0.0              0.0            0.000215
3               0.0              0.0            0.000242
4               0.0              0.0            0.000216

   btc_n_orphaned_blocks  btc_n_transactions_per_block  \
```

```
0                    0.0                           1.0
1                    0.0                           1.0
2                    0.0                           1.0
3                    0.0                           1.0
4                    0.0                           1.0

   btc_median_confirmation_time  ...  btc_cost_per_transaction_percent  \
0                           0.0  ...                          0.000000
1                           0.0  ...                         19.166667
2                           0.0  ...                          0.000000
3                           0.0  ...                        673.076923
4                           0.0  ...                          0.000000

   btc_cost_per_transaction  btc_n_unique_addresses  btc_n_transactions  \
0                       0.0                    71.0                71.0
1                       0.0                    71.0                78.0
2                       0.0                    70.0                70.0
3                       0.0                    73.0                73.0
4                       0.0                    83.0                83.0

   btc_n_transactions_total  btc_n_transactions_excluding_popular  \
0                   26958.0                                  71.0
1                   27036.0                                  78.0
2                   27106.0                                  70.0
3                   27179.0                                  73.0
4                   27262.0                                  83.0

   btc_n_transactions_excluding_chains_longer_than_100  btc_output_volume  \
0                                               71.0             3550.0
1                                               78.0            93450.0
2                                               70.0             3500.0
3                                               73.0             4100.0
4                                               83.0             4150.0

   btc_estimated_transaction_volume  btc_estimated_transaction_volume_usd
0                               0.0                                   0.0
1                           18000.0                                   0.0
2                               0.0                                   0.0
3                             520.0                                   0.0
4                               0.0                                   0.0

[5 rows x 24 columns]
Index(['Date', 'btc_market_price', 'btc_total_bitcoins', 'btc_market_cap',
       'btc_trade_volume', 'btc_blocks_size', 'btc_avg_block_size',
       'btc_n_orphaned_blocks', 'btc_n_transactions_per_block',
       'btc_median_confirmation_time', 'btc_hash_rate', 'btc_difficulty',
       'btc_miners_revenue', 'btc_transaction_fees',
       'btc_cost_per_transaction_percent', 'btc_cost_per_transaction',
       'btc_n_unique_addresses', 'btc_n_transactions',
       'btc_n_transactions_total', 'btc_n_transactions_excluding_popular',
       'btc_n_transactions_excluding_chains_longer_than_100',
       'btc_output_volume', 'btc_estimated_transaction_volume',
       'btc_estimated_transaction_volume_usd'],
      dtype='object')
```

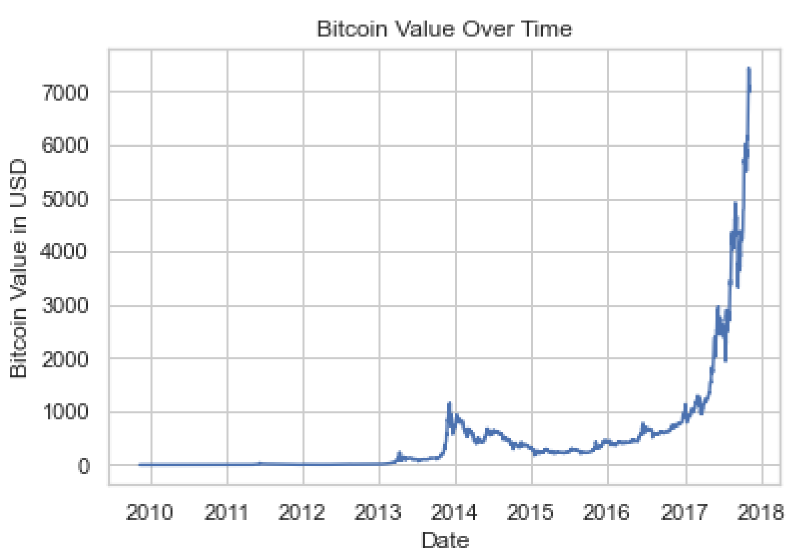
```
(2920, 24)
            Date  btc_market_price  btc_total_bitcoins  btc_market_cap  \
752   2011-12-02          3.138000           7787350.0    2.443670e+07
753   2011-12-03          3.129990           7794850.0    2.439780e+07
754   2011-12-04          2.990000           7801700.0    2.332708e+07
755   2011-12-05          2.930000           7809700.0    2.288242e+07
756   2011-12-06          3.050000           7817650.0    2.384383e+07
...          ...               ...                 ...             ...
2915  2017-11-03       7197.720060          16662275.0    1.199304e+11
2916  2017-11-04       7437.543317          16663900.0    1.239385e+11
2917  2017-11-05       7377.012367          16665662.5    1.229428e+11
2918  2017-11-06       6989.071667          16667325.0    1.164891e+11
2919  2017-11-07       7092.127233          16669275.0    1.182206e+11

      btc_trade_volume  btc_blocks_size  btc_avg_block_size  \
752       1.815046e+05       572.000000            0.017415
753       3.631256e+05       574.000000            0.016900
754       2.633752e+05       576.000000            0.015659
755       9.050023e+04       578.000000            0.017047
756       1.701647e+05       580.000000            0.018309
...                ...              ...                 ...
2915      3.748147e+08    139714.873251            1.046606
2916      5.635740e+08    139848.545492            1.028248
2917      5.685735e+08    139995.561562            1.042667
2918      8.328224e+08    140134.487161            1.044553
2919      5.339933e+08    140294.602454            1.026380

      btc_n_orphaned_blocks  btc_n_transactions_per_block  \
752                     0.0                     48.000000
753                     0.0                     37.000000
754                     0.0                     37.000000
755                     0.0                     36.000000
756                     0.0                     36.000000
...                     ...                           ...
2915                    0.0                   2332.756303
2916                    0.0                   2262.469231
2917                    0.0                   1785.304965
2918                    0.0                   2037.812030
2919                    0.0                   2151.512821

      btc_median_confirmation_time  ...  btc_cost_per_transaction  \
752                      11.066667  ...                  3.902130
753                      12.783333  ...                  4.244975
754                      16.016667  ...                  3.873435
755                      13.366667  ...                  3.694468
756                      15.183333  ...                  3.454871
...                            ...  ...                       ...
2915                     18.875000  ...                 45.615461
2916                     15.516667  ...                 48.001722
2917                     14.900000  ...                 57.443344
2918                     12.016667  ...                 49.068810
2919                     10.733333  ...                 47.419337

      btc_n_unique_addresses  btc_n_transactions  btc_n_transactions_total  \
```

```
752                10589.0              6316.0              1958524.0
753                 9426.0              5533.0              1964057.0
754                 9094.0              5290.0              1969347.0
755                10411.0              6347.0              1975694.0
756                11341.0              7021.0              1982715.0
...                    ...                 ...                    ...
2915              670928.0            277598.0            268308467.0
2916              588058.0            294121.0            268602588.0
2917              555458.0            251728.0            268854316.0
2918              608650.0            271029.0            269125345.0
2919              714349.0            335636.0            269460981.0

      btc_n_transactions_excluding_popular  \
752                                 4961.0
753                                 4169.0
754                                 3899.0
755                                 4689.0
756                                 5580.0
...                                    ...
2915                              268570.0
2916                              283626.0
2917                              243244.0
2918                              260465.0
2919                              323686.0

      btc_n_transactions_excluding_chains_longer_than_100  btc_output_volume  \
752                                              5779.0         4.672087e+06
753                                              5018.0         4.034196e+06
754                                              5170.0         3.364466e+06
755                                              5866.0         2.092724e+06
756                                              6302.0         7.369513e+06
...                                                 ...                  ...
2915                                           203967.0         3.036930e+06
2916                                           228764.0         5.765969e+06
2917                                           180241.0         6.077425e+06
2918                                           199667.0         2.782649e+06
2919                                           241986.0         3.523838e+06

      btc_estimated_transaction_volume  btc_estimated_transaction_volume_usd  \
752                       3.308930e+06                          1.038342e+07
753                       3.682124e+06                          1.152501e+07
754                       3.204551e+06                          9.581607e+06
755                       1.883357e+06                          5.518235e+06
756                       5.825066e+06                          1.776645e+07
...                                ...                                   ...
2915                      2.323145e+05                          1.672135e+09
2916                      1.915101e+05                          1.424365e+09
2917                      1.775416e+05                          1.309727e+09
2918                      2.244207e+05                          1.568492e+09
2919                      2.574699e+05                          1.826009e+09

      Days
752    752
753    753
```

```
754     754
755     755
756     756
...     ...
2915   2915
2916   2916
2917   2917
2918   2918
2919   2919

[2168 rows x 25 columns]
/Users/alvaroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/l
ib/python3.9/site-packages/pandas/util/_decorators.py:311: SettingWithCopyWarnin
g:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stab
le/user_guide/indexing.html#returning-a-view-versus-a-copy
  return func(*args, **kwargs)
```

In [5]:

```
# Plot btc_market_price vs Date with a scatter plot
plt.plot(BTC_data["Date"], BTC_data["btc_market_price"])
plt.title("Bitcoin Value Over Time")
plt.xlabel("Date")
plt.ylabel("Bitcoin Value in USD")
plt.show()
```



We can see that the value of bitcoin has grown exponentially.

Exploratory data analysis to data set

Since market fluctuations are hard to predict and we can only analyze observational data of the past, and model based on percent change will give an accurate picture of the correlation between the most significant variables affecting bitcoin's market price. Therefore, as price increases, so will the relative variations of the highest dependent features of Bitcoin's blockchain. This corroborates the initial hypothesis that bitcoin's price is dictated by the law of supply and demand. The reason for that is that eventually the percent change in the count of bitcoin will become irrelevant because of its finite supply of 21 million coins.

Correlation between variables

In [6]:

```
corr = BTC_data.corr()
corr = corr.iloc[1:4, 5:25].corr()
# Plot correlation matrix
sns.heatmap(corr, annot=True, xticklabels=corr.columns,
yticklabels=corr.columns)
print(corr)
```

```
                                                   btc_avg_block_size  \
btc_avg_block_size                                           1.000000
btc_n_orphaned_blocks                                        0.921531
btc_n_transactions_per_block                                 0.999818
btc_median_confirmation_time                                 0.979275
btc_hash_rate                                               -0.756919
btc_difficulty                                              -0.737205
btc_miners_revenue                                          -0.791462
btc_transaction_fees                                        -0.559594
btc_cost_per_transaction_percent                            -0.943949
btc_cost_per_transaction                                    -0.456912
btc_n_unique_addresses                                       0.996330
btc_n_transactions                                           0.999945
btc_n_transactions_total                                     0.749527
btc_n_transactions_excluding_popular                         0.999640
btc_n_transactions_excluding_chains_longer_than...           0.986272
btc_output_volume                                            0.921520
btc_estimated_transaction_volume                             0.876795
btc_estimated_transaction_volume_usd                        -0.870488
Days                                                         0.987685

                                                   btc_n_orphaned_blocks  \
btc_avg_block_size                                              0.921531
btc_n_orphaned_blocks                                           1.000000
btc_n_transactions_per_block                                    0.913945
btc_median_confirmation_time                                    0.981078
btc_hash_rate                                                  -0.951285
btc_difficulty                                                 -0.941723
btc_miners_revenue                                             -0.966696
btc_transaction_fees                                           -0.837498
btc_cost_per_transaction_percent                               -0.998054
btc_cost_per_transaction                                       -0.766460
btc_n_unique_addresses                                          0.884912
btc_n_transactions                                              0.925550
btc_n_transactions_total                                        0.433664
```

```
btc_n_transactions_excluding_popular                         0.910783
btc_n_transactions_excluding_chains_longer_than...           0.844759
btc_output_volume                                            1.000000
btc_estimated_transaction_volume                             0.994716
btc_estimated_transaction_volume_usd                        -0.993301
Days                                                         0.970935

                                                   btc_n_transactions_per_block
\
btc_avg_block_size                                                     0.999818
btc_n_orphaned_blocks                                                  0.913945
btc_n_transactions_per_block                                           1.000000
btc_median_confirmation_time                                           0.975227
btc_hash_rate                                                         -0.744296
btc_difficulty                                                        -0.724163
btc_miners_revenue                                                    -0.779641
btc_transaction_fees                                                  -0.543660
btc_cost_per_transaction_percent                                      -0.937471
btc_cost_per_transaction                                              -0.439836
btc_n_unique_addresses                                                 0.997783
btc_n_transactions                                                     0.999562
btc_n_transactions_total                                               0.762036
btc_n_transactions_excluding_popular                                   0.999970
btc_n_transactions_excluding_chains_longer_than...                     0.989246
btc_output_volume                                                      0.913933
btc_estimated_transaction_volume                                       0.867449
btc_estimated_transaction_volume_usd                                  -0.860927
Days                                                                   0.984516

                                                   btc_median_confirmation_time
\
btc_avg_block_size                                                     0.979275
btc_n_orphaned_blocks                                                  0.981078
btc_n_transactions_per_block                                           0.975227
btc_median_confirmation_time                                           1.000000
btc_hash_rate                                                         -0.873590
btc_difficulty                                                        -0.858773
btc_miners_revenue                                                    -0.898852
btc_transaction_fees                                                  -0.715851
btc_cost_per_transaction_percent                                      -0.991241
btc_cost_per_transaction                                              -0.627600
btc_n_unique_addresses                                                 0.958345
btc_n_transactions                                                     0.981344
btc_n_transactions_total                                               0.599919
btc_n_transactions_excluding_popular                                   0.973490
btc_n_transactions_excluding_chains_longer_than...                     0.932387
btc_output_volume                                                      0.981072
btc_estimated_transaction_volume                                       0.956015
btc_estimated_transaction_volume_usd                                  -0.952133
Days                                                                   0.998903

                                                   btc_hash_rate  \
btc_avg_block_size                                     -0.756919
btc_n_orphaned_blocks                                  -0.951285
```

```
btc_n_transactions_per_block                          -0.744296
btc_median_confirmation_time                          -0.873590
btc_hash_rate                                          1.000000
btc_difficulty                                         0.999560
btc_miners_revenue                                     0.998509
btc_transaction_fees                                   0.965174
btc_cost_per_transaction_percent                       0.930211
btc_cost_per_transaction                               0.927149
btc_n_unique_addresses                                -0.698203
btc_n_transactions                                    -0.763726
btc_n_transactions_total                              -0.134724
btc_n_transactions_excluding_popular                  -0.739116
btc_n_transactions_excluding_chains_longer_than...    -0.638614
btc_output_volume                                     -0.951294
btc_estimated_transaction_volume                      -0.977912
btc_estimated_transaction_volume_usd                   0.980539
Days                                                  -0.849842

                                                    btc_difficulty  \
btc_avg_block_size                                       -0.737205
btc_n_orphaned_blocks                                    -0.941723
btc_n_transactions_per_block                             -0.724163
btc_median_confirmation_time                             -0.858773
btc_hash_rate                                             0.999560
btc_difficulty                                            1.000000
btc_miners_revenue                                        0.996451
btc_transaction_fees                                      0.972508
btc_cost_per_transaction_percent                          0.918917
btc_cost_per_transaction                                  0.937854
btc_n_unique_addresses                                   -0.676665
btc_n_transactions                                       -0.744246
btc_n_transactions_total                                 -0.105279
btc_n_transactions_excluding_popular                     -0.718815
btc_n_transactions_excluding_chains_longer_than...       -0.615511
btc_output_volume                                        -0.941732
btc_estimated_transaction_volume                         -0.971283
btc_estimated_transaction_volume_usd                      0.974285
Days                                                     -0.833838

                                                    btc_miners_revenue  \
btc_avg_block_size                                           -0.791462
btc_n_orphaned_blocks                                        -0.966696
btc_n_transactions_per_block                                 -0.779641
btc_median_confirmation_time                                 -0.898852
btc_hash_rate                                                 0.998509
btc_difficulty                                                0.996451
btc_miners_revenue                                            1.000000
btc_transaction_fees                                          0.949455
btc_cost_per_transaction_percent                              0.948858
btc_cost_per_transaction                                      0.905315
btc_n_unique_addresses                                       -0.736239
btc_n_transactions                                           -0.797824
btc_n_transactions_total                                     -0.188610
btc_n_transactions_excluding_popular                         -0.774781
```

```
btc_n_transactions_excluding_chains_longer_than...          -0.679666
btc_output_volume                                           -0.966703
btc_estimated_transaction_volume                            -0.987863
btc_estimated_transaction_volume_usd                         0.989793
Days                                                        -0.877344

                                                    btc_transaction_fees  \
btc_avg_block_size                                             -0.559594
btc_n_orphaned_blocks                                          -0.837498
btc_n_transactions_per_block                                   -0.543660
btc_median_confirmation_time                                   -0.715851
btc_hash_rate                                                   0.965174
btc_difficulty                                                  0.972508
btc_miners_revenue                                              0.949455
btc_transaction_fees                                            1.000000
btc_cost_per_transaction_percent                                0.801798
btc_cost_per_transaction                                        0.992883
btc_n_unique_addresses                                         -0.486602
btc_n_transactions                                             -0.568250
btc_n_transactions_total                                        0.129191
btc_n_transactions_excluding_popular                           -0.537162
btc_n_transactions_excluding_chains_longer_than...             -0.415058
btc_output_volume                                              -0.837514
btc_estimated_transaction_volume                               -0.889174
btc_estimated_transaction_volume_usd                            0.895030
Days                                                           -0.682368

                                                    btc_cost_per_transaction_per
cent  \
btc_avg_block_size                                                         -0.94
3949
btc_n_orphaned_blocks                                                      -0.99
8054
btc_n_transactions_per_block                                               -0.93
7471
btc_median_confirmation_time                                               -0.99
1241
btc_hash_rate                                                               0.93
0211
btc_difficulty                                                              0.91
8917
btc_miners_revenue                                                          0.94
8858
btc_transaction_fees                                                        0.80
1798
btc_cost_per_transaction_percent                                            1.00
0000
btc_cost_per_transaction                                                    0.72
4922
btc_n_unique_addresses                                                     -0.91
2230
btc_n_transactions                                                         -0.94
7357
btc_n_transactions_total                                                   -0.48
```

```
9002
btc_n_transactions_excluding_popular                              -0.93
4755
btc_n_transactions_excluding_chains_longer_than...                -0.87
6482
btc_output_volume                                                 -0.99
8053
btc_estimated_transaction_volume                                  -0.98
6379
btc_estimated_transaction_volume_usd                               0.98
4164
Days                                                              -0.98
3969

                                                    btc_cost_per_transaction  \
btc_avg_block_size                                                 -0.456912
btc_n_orphaned_blocks                                              -0.766460
btc_n_transactions_per_block                                       -0.439836
btc_median_confirmation_time                                       -0.627600
btc_hash_rate                                                       0.927149
btc_difficulty                                                      0.937854
btc_miners_revenue                                                  0.905315
btc_transaction_fees                                                0.992883
btc_cost_per_transaction_percent                                    0.724922
btc_cost_per_transaction                                            1.000000
btc_n_unique_addresses                                             -0.379096
btc_n_transactions                                                 -0.466210
btc_n_transactions_total                                            0.246366
btc_n_transactions_excluding_popular                               -0.432887
btc_n_transactions_excluding_chains_longer_than...                 -0.303754
btc_output_volume                                                  -0.766479
btc_estimated_transaction_volume                                   -0.828353
btc_estimated_transaction_volume_usd                                0.835545
Days                                                               -0.590454

                                                    btc_n_unique_addresses  \
btc_avg_block_size                                                0.996330
btc_n_orphaned_blocks                                             0.884912
btc_n_transactions_per_block                                      0.997783
btc_median_confirmation_time                                      0.958345
btc_hash_rate                                                    -0.698203
btc_difficulty                                                   -0.676665
btc_miners_revenue                                               -0.736239
btc_transaction_fees                                             -0.486602
btc_cost_per_transaction_percent                                 -0.912230
btc_cost_per_transaction                                         -0.379096
btc_n_unique_addresses                                            1.000000
btc_n_transactions                                                0.995378
btc_n_transactions_total                                          0.803438
btc_n_transactions_excluding_popular                              0.998267
btc_n_transactions_excluding_chains_longer_than...                0.996787
btc_output_volume                                                 0.884898
btc_estimated_transaction_volume                                  0.832417
btc_estimated_transaction_volume_usd                             -0.825164
```

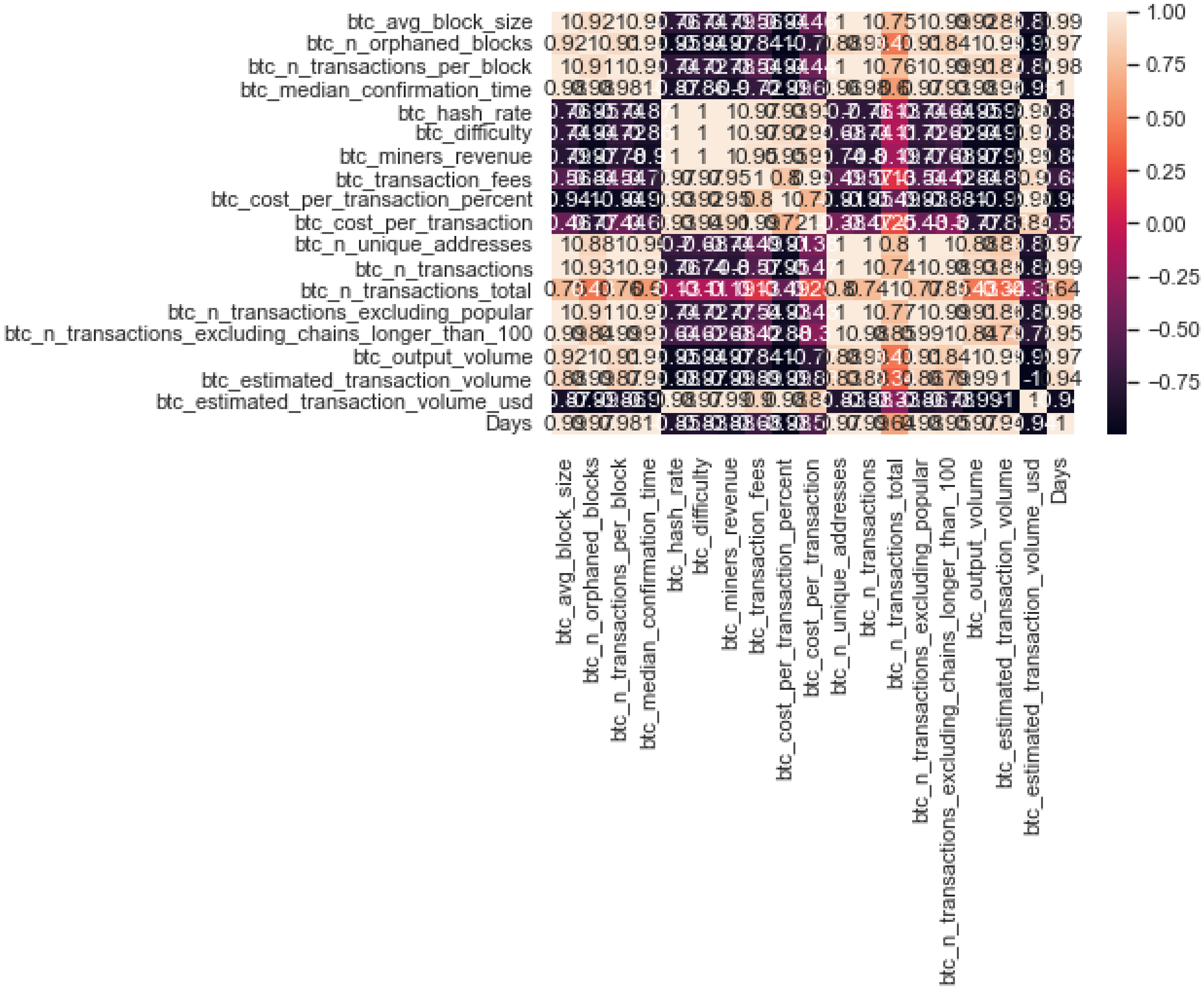
```
Days                                                               0.970668

                                                   btc_n_transactions  \
btc_avg_block_size                                           0.999945
btc_n_orphaned_blocks                                        0.925550
btc_n_transactions_per_block                                 0.999562
btc_median_confirmation_time                                 0.981344
btc_hash_rate                                               -0.763726
btc_difficulty                                              -0.744246
btc_miners_revenue                                          -0.797824
btc_transaction_fees                                        -0.568250
btc_cost_per_transaction_percent                            -0.947357
btc_cost_per_transaction                                    -0.466210
btc_n_unique_addresses                                       0.995378
btc_n_transactions                                           1.000000
btc_n_transactions_total                                     0.742547
btc_n_transactions_excluding_popular                         0.999304
btc_n_transactions_excluding_chains_longer_than...           0.984487
btc_output_volume                                            0.925539
btc_estimated_transaction_volume                             0.881786
btc_estimated_transaction_volume_usd                        -0.875599
Days                                                         0.989270

                                                   btc_n_transactions_total  \
btc_avg_block_size                                                 0.749527
btc_n_orphaned_blocks                                              0.433664
btc_n_transactions_per_block                                       0.762036
btc_median_confirmation_time                                       0.599919
btc_hash_rate                                                     -0.134724
btc_difficulty                                                    -0.105279
btc_miners_revenue                                                -0.188610
btc_transaction_fees                                               0.129191
btc_cost_per_transaction_percent                                  -0.489002
btc_cost_per_transaction                                           0.246366
btc_n_unique_addresses                                             0.803438
btc_n_transactions                                                 0.742547
btc_n_transactions_total                                           1.000000
btc_n_transactions_excluding_popular                               0.767014
btc_n_transactions_excluding_chains_longer_than...                 0.848549
btc_output_volume                                                  0.433638
btc_estimated_transaction_volume                                   0.338861
btc_estimated_transaction_volume_usd                              -0.326637
Days                                                               0.636726

                                                   btc_n_transactions_excluding
_popular  \
btc_avg_block_size
0.999640
btc_n_orphaned_blocks
0.910783
btc_n_transactions_per_block
0.999970
btc_median_confirmation_time
0.973490
```

```
btc_hash_rate                                                            -
0.739116
btc_difficulty                                                           -
0.718815
btc_miners_revenue                                                       -
0.774781
btc_transaction_fees                                                     -
0.537162
btc_cost_per_transaction_percent                                         -
0.934755
btc_cost_per_transaction                                                 -
0.432887
btc_n_unique_addresses
0.998267
btc_n_transactions
0.999304
btc_n_transactions_total
0.767014
btc_n_transactions_excluding_popular
1.000000
btc_n_transactions_excluding_chains_longer_than...
0.990346
btc_output_volume
0.910771
btc_estimated_transaction_volume
0.863580
btc_estimated_transaction_volume_usd                                     -
0.856972
Days
0.983133

                                                   btc_n_transactions_excluding
_chains_longer_than_100  \
btc_avg_block_size
0.986272
btc_n_orphaned_blocks
0.844759
btc_n_transactions_per_block
0.989246
btc_median_confirmation_time
0.932387
btc_hash_rate
-0.638614
btc_difficulty
-0.615511
btc_miners_revenue
-0.679666
btc_transaction_fees
-0.415058
btc_cost_per_transaction_percent
-0.876482
btc_cost_per_transaction
-0.303754
btc_n_unique_addresses
```

```
0.996787
btc_n_transactions
0.984487
btc_n_transactions_total
0.848549
btc_n_transactions_excluding_popular
0.990346
btc_n_transactions_excluding_chains_longer_than...
1.000000
btc_output_volume
0.844744
btc_estimated_transaction_volume
0.785353
btc_estimated_transaction_volume_usd
-0.777263
Days
0.948290

                                                    btc_output_volume  \
btc_avg_block_size                                           0.921520
btc_n_orphaned_blocks                                        1.000000
btc_n_transactions_per_block                                 0.913933
btc_median_confirmation_time                                 0.981072
btc_hash_rate                                               -0.951294
btc_difficulty                                              -0.941732
btc_miners_revenue                                          -0.966703
btc_transaction_fees                                        -0.837514
btc_cost_per_transaction_percent                            -0.998053
btc_cost_per_transaction                                    -0.766479
btc_n_unique_addresses                                       0.884898
btc_n_transactions                                           0.925539
btc_n_transactions_total                                     0.433638
btc_n_transactions_excluding_popular                         0.910771
btc_n_transactions_excluding_chains_longer_than...           0.844744
btc_output_volume                                            1.000000
btc_estimated_transaction_volume                             0.994719
btc_estimated_transaction_volume_usd                        -0.993305
Days                                                         0.970928

                                                    btc_estimated_transaction_vo
lume  \
btc_avg_block_size                                                          0.87
6795
btc_n_orphaned_blocks                                                       0.99
4716
btc_n_transactions_per_block                                                0.86
7449
btc_median_confirmation_time                                                0.95
6015
btc_hash_rate                                                              -0.97
7912
btc_difficulty                                                             -0.97
1283
btc_miners_revenue                                                         -0.98
```

```
7863
btc_transaction_fees                                          -0.88
9174
btc_cost_per_transaction_percent                              -0.98
6379
btc_cost_per_transaction                                      -0.82
8353
btc_n_unique_addresses                                         0.83
2417
btc_n_transactions                                             0.88
1786
btc_n_transactions_total                                       0.33
8861
btc_n_transactions_excluding_popular                           0.86
3580
btc_n_transactions_excluding_chains_longer_than...             0.78
5353
btc_output_volume                                              0.99
4719
btc_estimated_transaction_volume                               1.00
0000
btc_estimated_transaction_volume_usd                          -0.99
9916
Days                                                           0.94
1231

                                  btc_estimated_transaction_vo
lume_usd  \
btc_avg_block_size                                            -
0.870488
btc_n_orphaned_blocks                                         -
0.993301
btc_n_transactions_per_block                                  -
0.860927
btc_median_confirmation_time                                  -
0.952133
btc_hash_rate
0.980539
btc_difficulty
0.974285
btc_miners_revenue
0.989793
btc_transaction_fees
0.895030
btc_cost_per_transaction_percent
0.984164
btc_cost_per_transaction
0.835545
btc_n_unique_addresses                                        -
0.825164
btc_n_transactions                                            -
0.875599
btc_n_transactions_total                                      -
0.326637
```

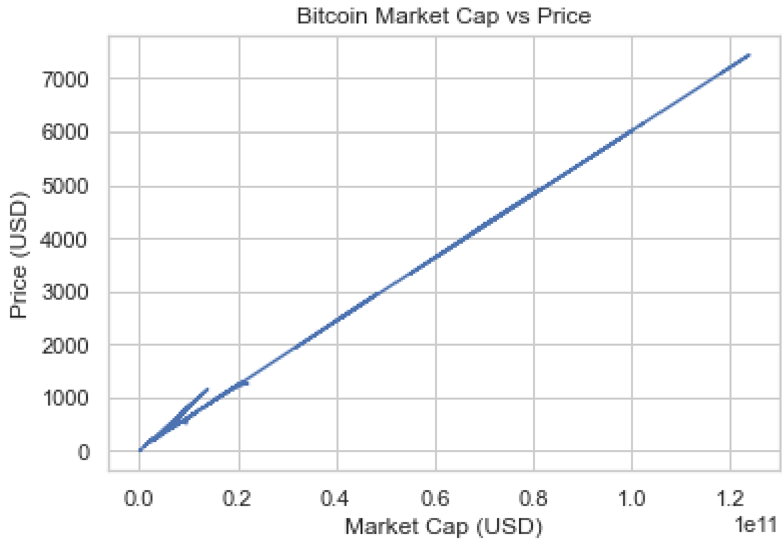
```
btc_n_transactions_excluding_popular                              -
0.856972
btc_n_transactions_excluding_chains_longer_than...                -
0.777263
btc_output_volume                                                 -
0.993305
btc_estimated_transaction_volume                                  -
0.999916
btc_estimated_transaction_volume_usd
1.000000
Days                                                              -
0.936774

                                                        Days
btc_avg_block_size                                  0.987685
btc_n_orphaned_blocks                               0.970935
btc_n_transactions_per_block                        0.984516
btc_median_confirmation_time                        0.998903
btc_hash_rate                                      -0.849842
btc_difficulty                                     -0.833838
btc_miners_revenue                                 -0.877344
btc_transaction_fees                               -0.682368
btc_cost_per_transaction_percent                   -0.983969
btc_cost_per_transaction                           -0.590454
btc_n_unique_addresses                              0.970668
btc_n_transactions                                  0.989270
btc_n_transactions_total                            0.636726
btc_n_transactions_excluding_popular                0.983133
btc_n_transactions_excluding_chains_longer_than...  0.948290
btc_output_volume                                   0.970928
btc_estimated_transaction_volume                    0.941231
btc_estimated_transaction_volume_usd               -0.936774
Days                                                1.000000
```



The most correlated variables to Market Price are Market Cap, Hash Rate, Difficulty, Miner Revenue, and Estimated USD Transaction Volume

Market Cap and Market Price

In [7]:
```
# Plot btc_market_cap vs btc_market_price with a scatter plot
plt.plot(BTC_data2["btc_market_cap"],
BTC_data2["btc_market_price"])
# Market Cap vs Price
plt.title("Bitcoin Market Cap vs Price")
plt.xlabel("Market Cap (USD)")
plt.ylabel("Price (USD)")
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

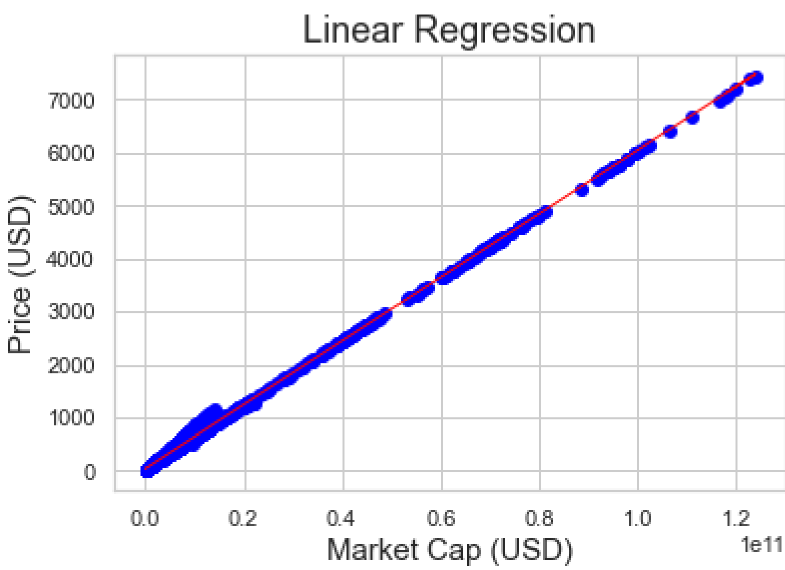


In [8]:

```
# Create a linear regression object
fit1 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
# Train the model using the training sets
fit1.fit(BTC_data2[["btc_market_cap"]],
BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions = fit1.predict(BTC_data2[["btc_market_cap"]])
# Residuals
residuals = BTC_data2["btc_market_price"] - predictions
# The coefficients
print(f'Coefficients: {fit1.coef_}')
# Intercept
print(f'Intercept: {fit1.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
fit1.score(BTC_data2[["btc_market_cap"]],
BTC_data2["btc_market_price"]))
# Plot outputs
plt.scatter(BTC_data2["btc_market_cap"],
BTC_data2["btc_market_price"],  color='blue')
```

```
plt.plot(BTC_data2["btc_market_cap"], predictions, color='red',
linewidth=1)
plt.title('Linear Regression')
plt.xlabel('Market Cap (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

```
Coefficients: [5.99422357e-08]
Intercept: 46.039960391727504
Mean squared error: 2447.80
Variance score: 1.00
```

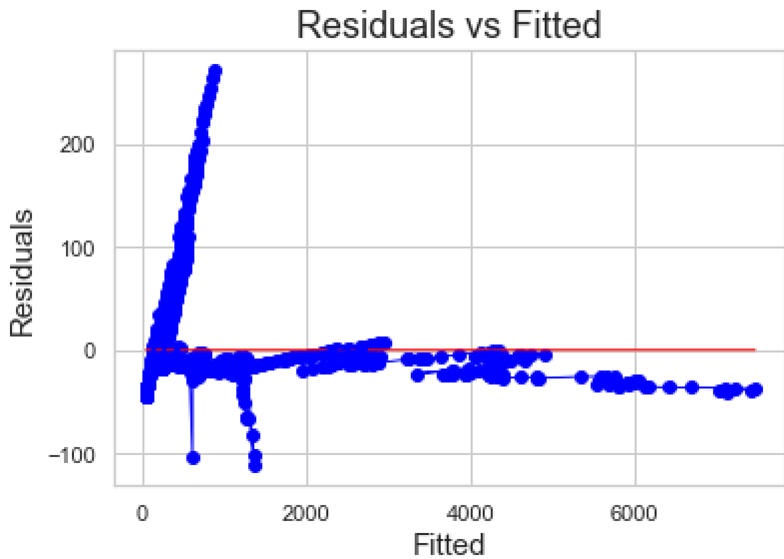


Since R-Squared is very close to 1, Market Capitalization is signficant to Market Price.

In [9]:

```
# Residuals vs fitted plot
plt.scatter(predictions, residuals,  color='blue')
plt.xticks(np.arange(0, max(predictions), step=2000))
plt.yticks(np.arange(-100, 300, step=100))
plt.plot(np.unique(predictions),
np.poly1d(np.polyfit(predictions, residuals, 1))
(np.unique(predictions)), color='red', linewidth=1)
plt.plot(predictions, residuals, color='blue', linewidth=1)
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
```

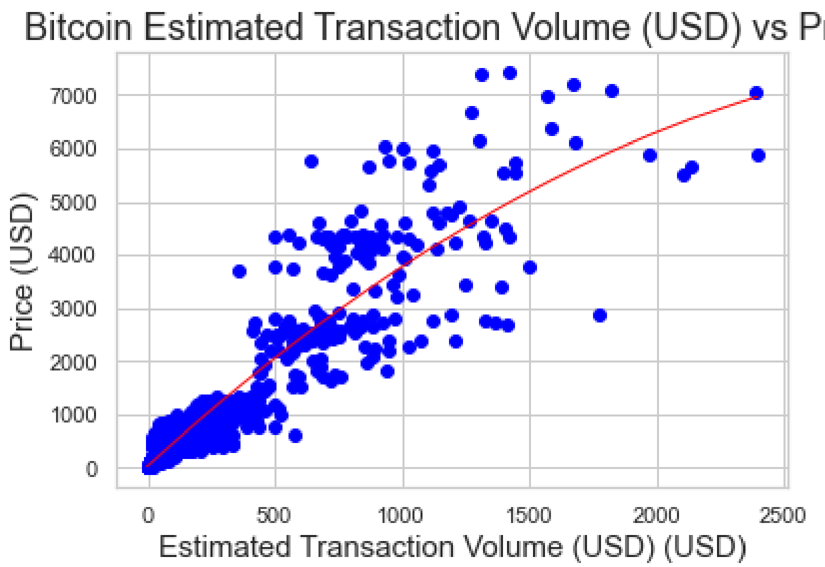
```
plt.ylabel('Residuals')
plt.show()
```



Estimated USD Transaction Volume vs. Market Price

In [10]:

```
# Plot btc_estimated_transaction_volume_usd vs btc_market_price
with a scatter plot
plt.plot(BTC_data2["btc_estimated_transaction_volume_usd"]/100000
 BTC_data2["btc_market_price"], 'o', color='blue', linewidth=1)
plt.plot(np.unique(BTC_data2["btc_estimated_transaction_volume_us

np.poly1d(np.polyfit(BTC_data2["btc_estimated_transaction_volume_
 BTC_data2["btc_market_price"], 2))
(np.unique(BTC_data2["btc_estimated_transaction_volume_usd"]/1000
 color='red', linewidth=1)
# Estimated Transaction Volume (USD) vs Price
plt.title("Bitcoin Estimated Transaction Volume (USD) vs
Price")
plt.xlabel("Estimated Transaction Volume (USD) (USD)")
plt.ylabel("Price (USD)")
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

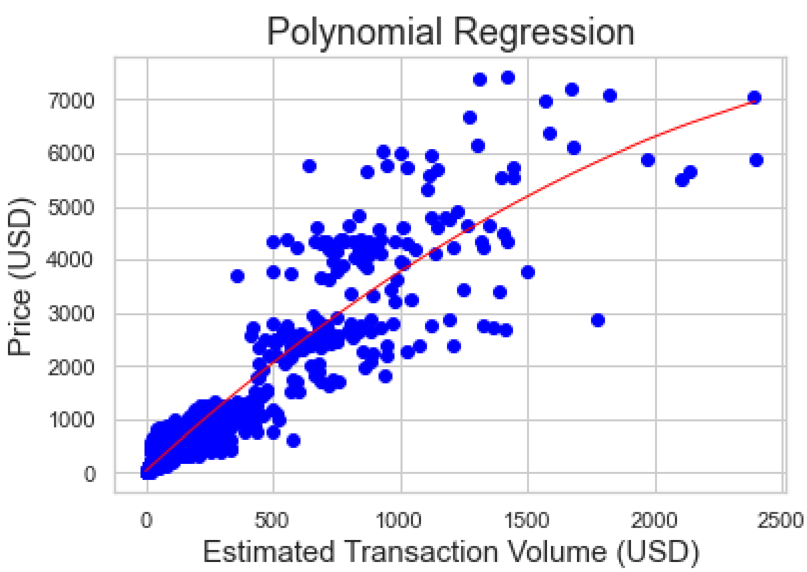


In [11]:

```
# Fit polynomial regression to btc_estimated_transaction_volume_usd squared vs btc_market_price
poly_btc_estimated_transaction_volume = PolynomialFeatures(degree=2).fit_transform(BTC_data2[["btc_estima

fit2 = linear_model.LinearRegression(fit_intercept=True, copy_X=True, n_jobs=1)
fit2.fit(poly_btc_estimated_transaction_volume, BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions2 = fit2.predict(poly_btc_estimated_transaction_volume)
# Residuals
residuals2 = BTC_data2["btc_market_price"] - predictions2
# The coefficients
print(f'Coefficients: {fit2.coef_}')
# Intercept
print(f'Intercept: {fit2.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions2 - BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
```

```
fit2.score(poly_btc_estimated_transaction_volume,
BTC_data2["btc_market_price"]))
# Plot outputs
plt.scatter(BTC_data2["btc_estimated_transaction_volume_usd"]/100
 BTC_data2["btc_market_price"],  color='blue')
plt.plot(np.unique(BTC_data2["btc_estimated_transaction_volume_us

np.poly1d(np.polyfit(BTC_data2["btc_estimated_transaction_volume_
 BTC_data2["btc_market_price"], 2))
(np.unique(BTC_data2["btc_estimated_transaction_volume_usd"]/1000
 color='red', linewidth=1)
plt.title('Polynomial Regression')
plt.xlabel('Estimated Transaction Volume (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

```
Coefficients: [ 0.00000000e+00  4.34114978e-06 -6.04679033e-16]
Intercept: 34.44178114396766
Mean squared error: 142605.89
Variance score: 0.87
```



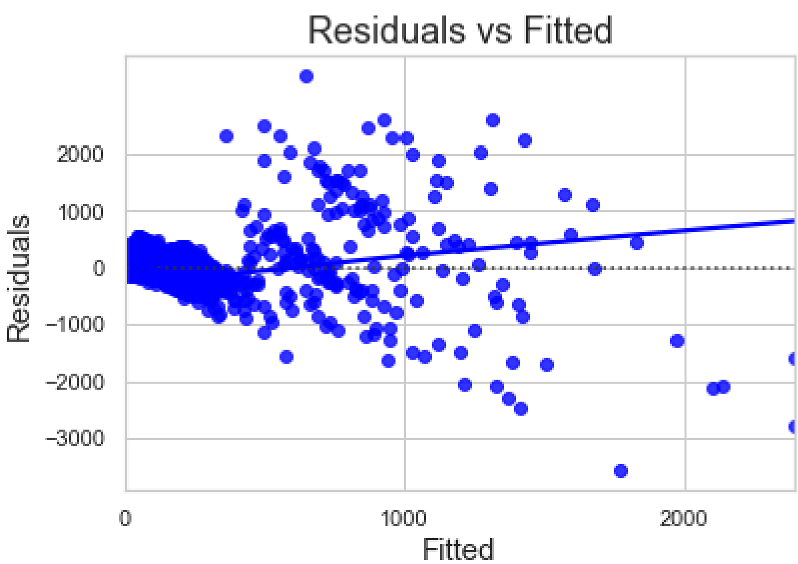
Transaction Volume is significant to Market Price due to R^2 = 0.86

In [12]:

```
# Residuals vs fitted plot for transaction volume
```

```
sns.residplot(BTC_data2["btc_estimated_transaction_volume_usd"]/1
 BTC_data2["btc_market_price"], lowess=True, color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.xticks(np.arange(0,
max(BTC_data2["btc_estimated_transaction_volume_usd"]/1000000),
step=1000))
plt.yticks(np.arange(-3000, 3000, step=1000))
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

```
/Users/alvaroserranorivas/.pyenv/versions/3.9.2/envs/bitcoin_linear_regression/l
ib/python3.9/site-packages/seaborn/_decorators.py:36: FutureWarning: Pass the fo
llowing variables as keyword args: x, y. From version 0.12, the only valid posit
ional argument will be `data`, and passing other arguments without an explicit k
eyword will result in an error or misinterpretation.
  warnings.warn(
```

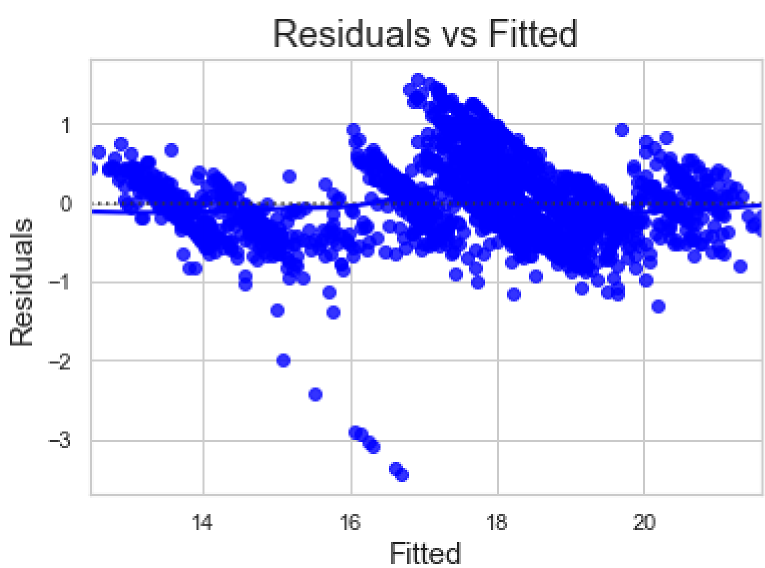


- Heteroscedasticity in the residuals plot is not ideal.
  - We can see that the residuals variance increases as the prediction values increase. As the price increases the variability increases. Therefore, a transformation of a variable in the model might be required.
  - Since most of the load appears to be in bottom, a log transformation will be attempted first.

Second iteration of Estimated Transaction Volume (USD) and market price

In [13]:

```
# Residuals vs fitted plot for linear model fit2b
sns.residplot(x=np.log(BTC_data2["btc_estimated_transaction_volum
 y=np.log(BTC_data2["btc_market_price"]), lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



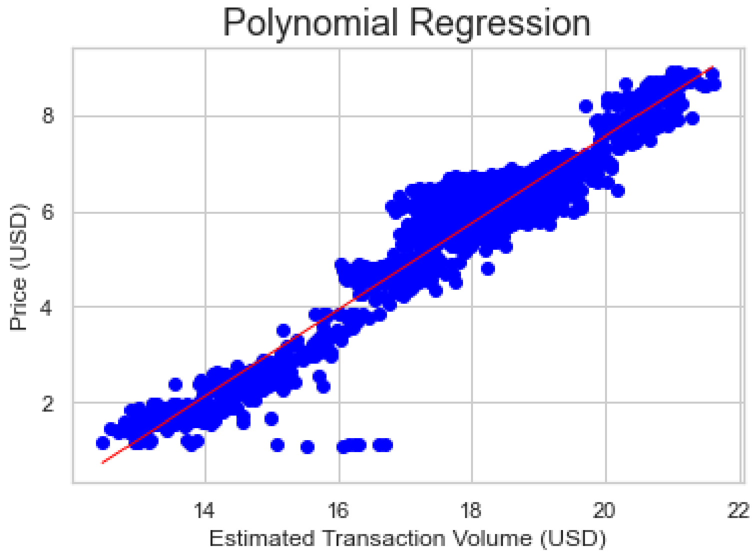
In [14]:

```
# Fit linear model for log transformed
btc_estimated_transaction_volume_usd vs log transformed
btc_market_price
fit2b = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit2b.fit(np.log(BTC_data2[["btc_estimated_transaction_volume_usd
 np.log(BTC_data2["btc_market_price"]))
# Make predictions using the testing set
predictions2b =
```

```
fit2b.predict(np.log(BTC_data2[["btc_estimated_transaction_volume

# Residuals
residuals2b = np.log(BTC_data2["btc_market_price"]) -
np.log(predictions2b)
# The coefficients
print(f'Coefficients: {fit2b.coef_}')
# Intercept
print(f'Intercept: {fit2b.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions2b -
np.log(BTC_data2["btc_market_price"])) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
fit2b.score(np.log(BTC_data2[["btc_estimated_transaction_volume_u
 np.log(BTC_data2["btc_market_price"])))
# Plot outputs
plt.scatter(np.log(BTC_data2["btc_estimated_transaction_volume_us
 np.log(BTC_data2["btc_market_price"]),  color='blue')
plt.plot(np.unique(np.log(BTC_data2["btc_estimated_transaction_vo

np.poly1d(np.polyfit(np.log(BTC_data2["btc_estimated_transaction_
 np.log(BTC_data2["btc_market_price"]), 1))
(np.unique(np.log(BTC_data2["btc_estimated_transaction_volume_usd
 color='red', linewidth=1)
plt.title('Polynomial Regression')
plt.xlabel('Estimated Transaction Volume (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

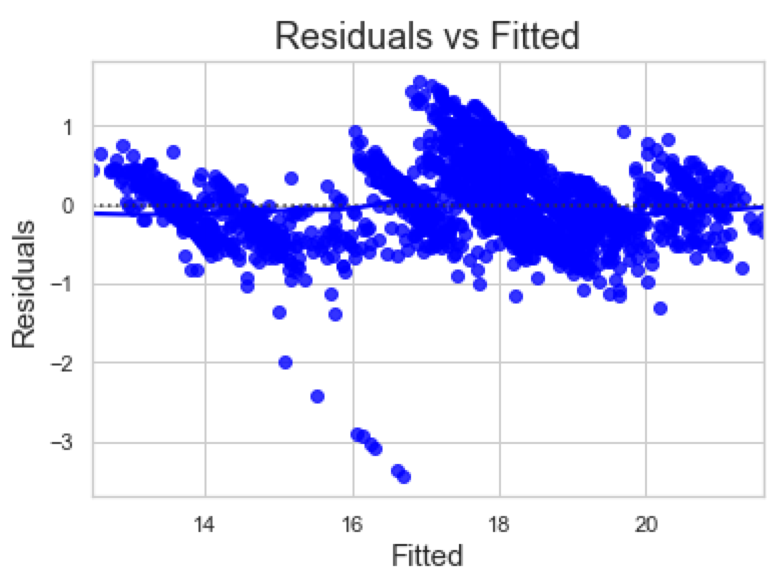
```
Coefficients: [0.90880364]
Intercept: -10.605824739229279
```

```
Mean squared error: 0.23
Variance score: 0.93
```



In [15]:

```
sns.residplot(x=np.log(BTC_data2["btc_estimated_transaction_volum
 y=np.log(BTC_data2["btc_market_price"]), lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



- The log transformation improves the heteroscedasticity issue significantly in the x direction
- The dispersion on the y-axis is not ideal but less of pattern

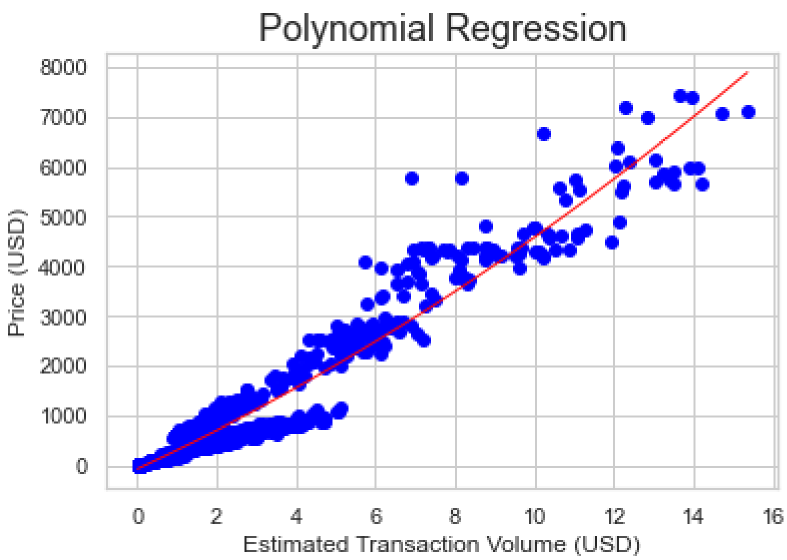
Miners Revenue and market price

In [16]:

```
# Fit polynomial regression for btc_miners_revenue vs
btc_market_price
poly_btc_miners_revenue =
PolynomialFeatures(degree=2).fit_transform(BTC_data2[["btc_miners

fit3 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit3.fit(poly_btc_miners_revenue,
BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions3 = fit3.predict(poly_btc_miners_revenue)
# Residuals
residuals3 = BTC_data2["btc_market_price"] - predictions3
# The coefficients
print(f'Coefficients: {fit3.coef_}')
# Intercept
print(f'Intercept: {fit3.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions3 -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
fit3.score(poly_btc_miners_revenue,
BTC_data2["btc_market_price"]))
# Plot outputs
plt.scatter(BTC_data2["btc_miners_revenue"]/1000000,
BTC_data2["btc_market_price"],  color='blue')
plt.plot(np.unique(BTC_data2["btc_miners_revenue"]/1000000),
np.poly1d(np.polyfit(BTC_data2["btc_miners_revenue"]/1000000,
BTC_data2["btc_market_price"], 2))
(np.unique(BTC_data2["btc_miners_revenue"]/1000000)),
```

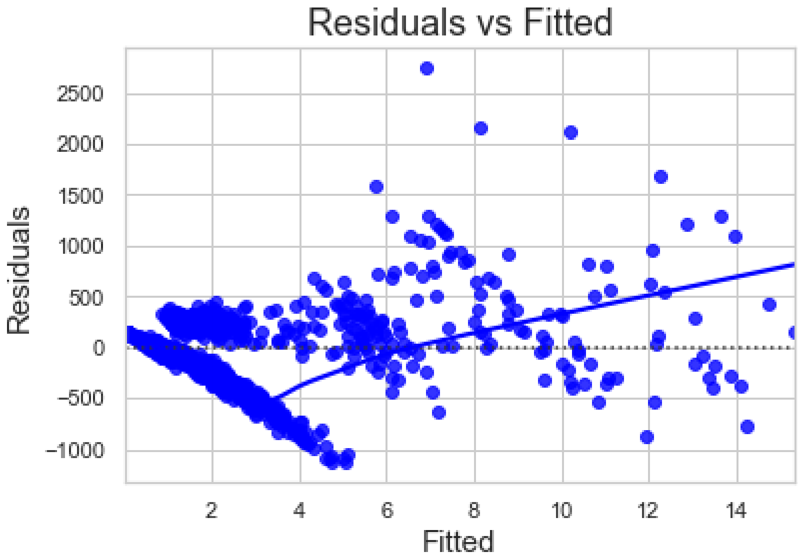
```
color='red', linewidth=1)
plt.title('Polynomial Regression')
plt.xlabel('Estimated Transaction Volume (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

```
Coefficients: [0.00000000e+00 3.65309044e-04 9.94627427e-12]
Intercept: -67.27876218906135
Mean squared error: 75171.20
Variance score: 0.93
```



In [17]:

```
# Residuals vs Fitted
sns.residplot(x=BTC_data2["btc_miners_revenue"]/1000000,
y=BTC_data2["btc_market_price"], lowess=True, color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



We can see that residuals are clustered in a certain area and do not show the best dispersion in terms of heteroscedasticity

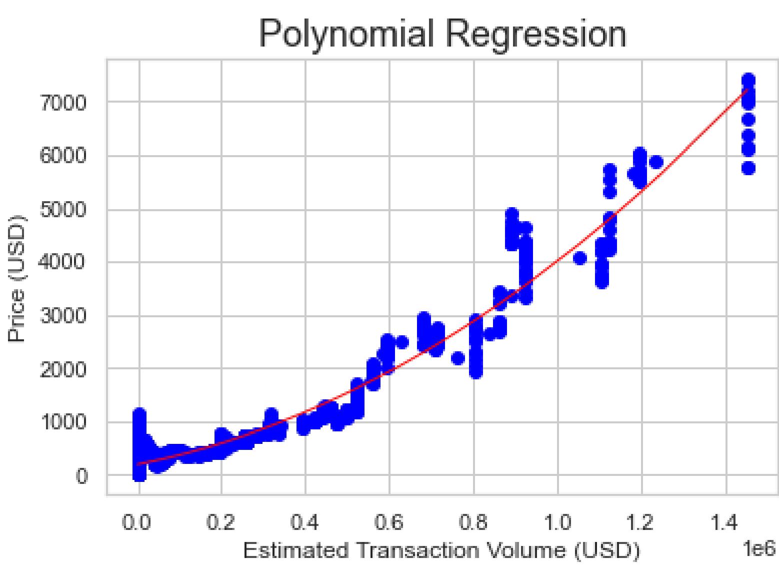
Difficulty and Market Price

In [18]:

```
# Fit polynomial regression for btc_difficulty vs
btc_market_price
poly_btc_difficulty =
PolynomialFeatures(degree=2).fit_transform(BTC_data2[["btc_diffic

fit4 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit4.fit(poly_btc_difficulty, BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions4 = fit4.predict(poly_btc_difficulty)
# Residuals
residuals4 = BTC_data2["btc_market_price"] - predictions4
# The coefficients
print(f'Coefficients: {fit4.coef_}')
# Intercept
print(f'Intercept: {fit4.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions4 -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
```

```
print('Variance score: %.2f' % fit4.score(poly_btc_difficulty,
BTC_data2["btc_market_price"]))
# Plot outputs
plt.scatter(BTC_data2["btc_difficulty"]/1000000,
BTC_data2["btc_market_price"],  color='blue')
plt.plot(np.unique(BTC_data2["btc_difficulty"]/1000000),
np.poly1d(np.polyfit(BTC_data2["btc_difficulty"]/1000000,
BTC_data2["btc_market_price"], 2))
(np.unique(BTC_data2["btc_difficulty"]/1000000)), color='red',
linewidth=1)
plt.title('Polynomial Regression')
plt.xlabel('Estimated Transaction Volume (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

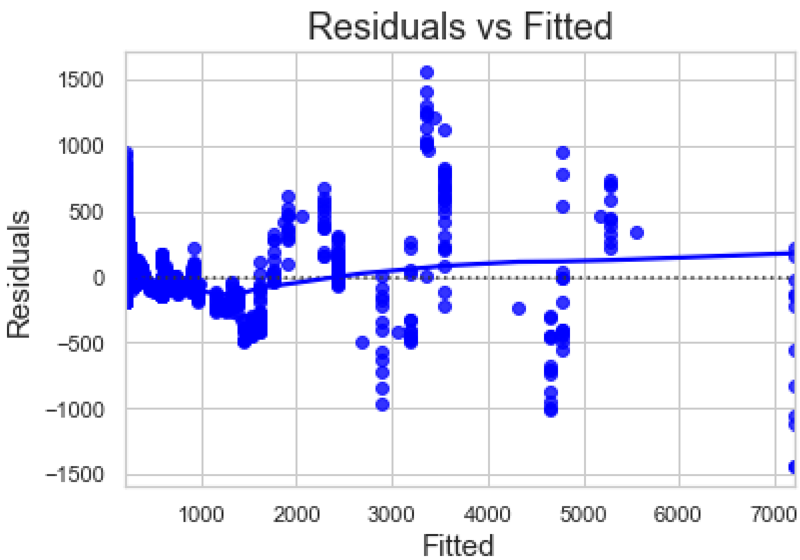
```
Coefficients: [0.00000000e+00 1.52149420e-09 2.27722823e-21]
Intercept: 199.6913915205463
Mean squared error: 69117.55
Variance score: 0.94
```



In [19]:

```
# Plot predictions4 vs residuals4 with sns
sns.residplot(x=predictions4, y=residuals4, lowess=True,
color='blue')
```

```
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



We can see some patterns, and therefore, reject a random dispersion

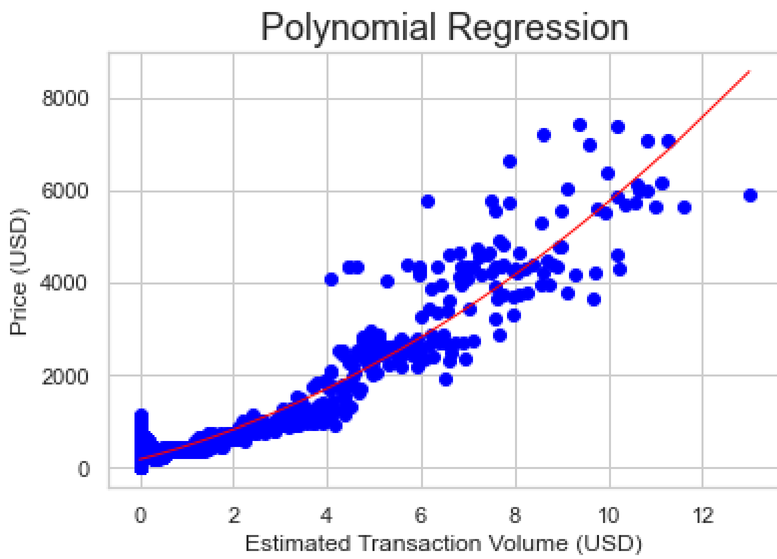
Hash Rate and Market Price

In [20]:

```
# Fit polynomial regression for btc_hash_rate vs
btc_market_price
poly_btc_hash_rate =
PolynomialFeatures(degree=2).fit_transform(BTC_data2[["btc_hash_r

fit5 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit5.fit(poly_btc_hash_rate, BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions5 = fit5.predict(poly_btc_hash_rate)
# Residuals
residuals5 = BTC_data2["btc_market_price"] - predictions5
# The coefficients
```

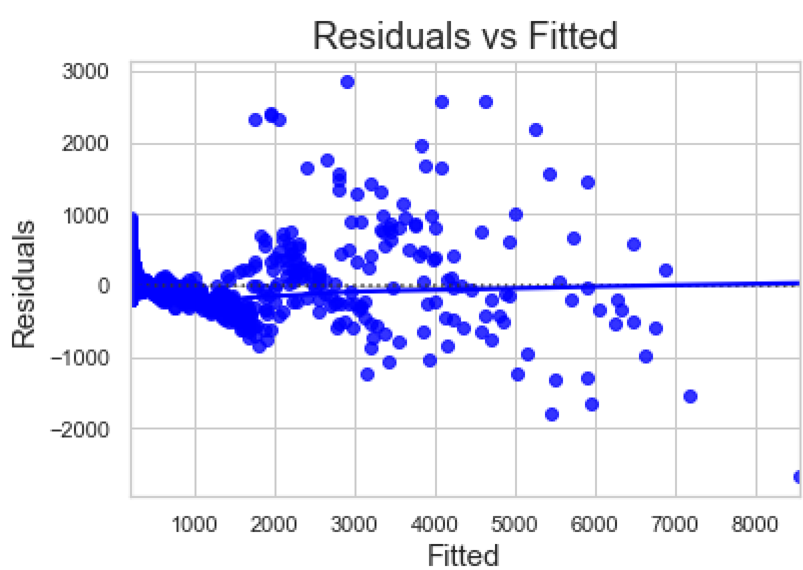
```
print(f'Coefficients: {fit5.coef_}')
# Intercept
print(f'Intercept: {fit5.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions5 -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' % fit5.score(poly_btc_hash_rate,
BTC_data2["btc_market_price"]))
# Plot outputs
plt.scatter(BTC_data2["btc_hash_rate"]/1000000,
BTC_data2["btc_market_price"],  color='blue')
plt.plot(np.unique(BTC_data2["btc_hash_rate"]/1000000),
np.poly1d(np.polyfit(BTC_data2["btc_hash_rate"]/1000000,
BTC_data2["btc_market_price"], 2))
(np.unique(BTC_data2["btc_hash_rate"]/1000000)), color='red',
linewidth=1)
plt.title('Polynomial Regression')
plt.xlabel('Estimated Transaction Volume (USD)')
plt.ylabel('Price (USD)')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=15)
plt.rc('axes', titlesize=18)
plt.show()
```

```
Coefficients: [0.00000000e+00 2.63592983e-04 2.93199784e-11]
Intercept: 184.0088987538802
Mean squared error: 110833.67
Variance score: 0.90
```



In [21]:

```
# Residuals vs Fitted
sns.residplot(x=predictions5, y=residuals5, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



Despite some heteroscedasticity the residuals mostly follow a flat line

Significance Analysis of All Variables to Market Price

In [22]:
```
print(BTC_data2.columns.values)
```

```
['Date' 'btc_market_price' 'btc_total_bitcoins' 'btc_market_cap'
 'btc_trade_volume' 'btc_blocks_size' 'btc_avg_block_size'
 'btc_n_orphaned_blocks' 'btc_n_transactions_per_block'
 'btc_median_confirmation_time' 'btc_hash_rate' 'btc_difficulty'
 'btc_miners_revenue' 'btc_transaction_fees'
 'btc_cost_per_transaction_percent' 'btc_cost_per_transaction'
 'btc_n_unique_addresses' 'btc_n_transactions' 'btc_n_transactions_total'
 'btc_n_transactions_excluding_popular'
 'btc_n_transactions_excluding_chains_longer_than_100' 'btc_output_volume'
 'btc_estimated_transaction_volume' 'btc_estimated_transaction_volume_usd'
 'Days']
```

In [23]:
```
# Run linear regression on all variables vs btc_market_price
fit6 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit6.fit(BTC_data2[["btc_difficulty", "btc_hash_rate",
"btc_market_cap", "btc_estimated_transaction_volume_usd",
"btc_output_volume","btc_n_transactions_total"
,"btc_trade_volume"]], BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions6 = fit6.predict(BTC_data2[["btc_difficulty",
"btc_hash_rate", "btc_market_cap",
"btc_estimated_transaction_volume_usd",
"btc_output_volume","btc_n_transactions_total"
,"btc_trade_volume"]])
# Residuals
residuals6 = BTC_data2["btc_market_price"] - predictions6
# The coefficients
print(f'Coefficients: {fit6.coef_}')
# Intercept
print(f'Intercept: {fit6.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions6 -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
```

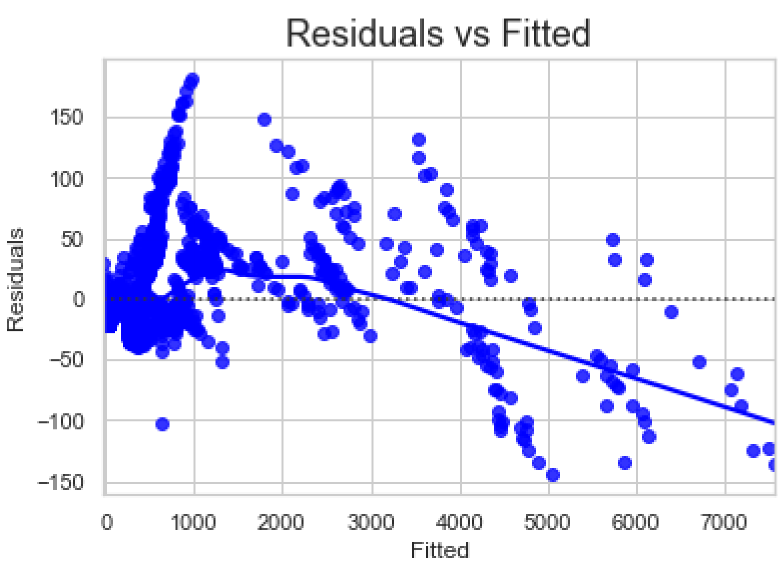
```
fit6.score(BTC_data2[["btc_difficulty", "btc_hash_rate",
"btc_market_cap", "btc_estimated_transaction_volume_usd",
"btc_output_volume","btc_n_transactions_total"
,"btc_trade_volume"]], BTC_data2["btc_market_price"]))
```

```
Coefficients: [-6.09934168e-10 -7.11124066e-06  6.73424241e-08  1.78257397e-09
 -1.47371627e-06  7.09254435e-07 -4.21248993e-08]
Intercept: 19.050210647484846
Mean squared error: 1319.51
Variance score: 1.00
```

R^2 score is very close to 1

In [24]:

```
# Residuals vs Fitted
sns.residplot(x=predictions6, y=residuals6, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



Even though most of the volume is closer to small values on the x-axis, most of the trend line is relatively flat.

Highly Correlated Variable vs Market Price

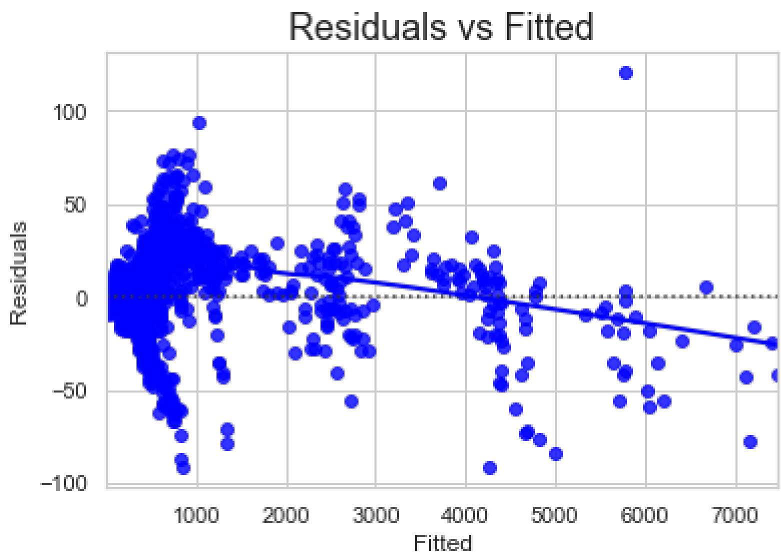
In [25]:
```
# Linear model of Highly correlated variables vs Market Price
fit7 = linear_model.LinearRegression(fit_intercept=True,
copy_X=True, n_jobs=1)
fit7.fit(BTC_data2[["btc_difficulty", "btc_hash_rate",
"btc_market_cap", "btc_estimated_transaction_volume_usd",
"btc_miners_revenue"]], BTC_data2["btc_market_price"])
# Make predictions using the testing set
predictions7 = fit7.predict(BTC_data2[["btc_difficulty",
"btc_hash_rate", "btc_market_cap",
"btc_estimated_transaction_volume_usd", "btc_miners_revenue"]])
# Residuals
residuals7 = BTC_data2["btc_market_price"] - predictions7
# The coefficients
print(f'Coefficients: {fit7.coef_}')
# Intercept
print(f'Intercept: {fit7.intercept_}')
# The mean squared error
print("Mean squared error: %.2f"
          % np.mean((predictions7 -
BTC_data2["btc_market_price"]) ** 2))
# Explained variance score: 1 is perfect prediction
print('Variance score: %.2f' %
fit7.score(BTC_data2[["btc_difficulty", "btc_hash_rate",
"btc_market_cap", "btc_estimated_transaction_volume_usd",
"btc_miners_revenue"]], BTC_data2["btc_market_price"]))
```

```
Coefficients: [ 4.62077875e-10 -6.36732046e-05  5.21760429e-08 -3.07034342e-08
  7.14661614e-05]
Intercept: 7.08707499472996
Mean squared error: 470.90
Variance score: 1.00
```

It appears that all of the highly correlated vairables to Market Price (Market Cap, Hash Rate, BTC Difficulty, Miners Revenue, and Estimated Transaction Volume USD) are significant.

In [26]:
```
# Residuals vs Fitted
sns.residplot(x=predictions7, y=residuals7, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
```

```
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



We do not see neither a dispersion nor a trend line, so a better adjustment is needed

Narrow down variables

- Market Capitalization and Estimated Transaction Volume are highly correlated, only one will be included in the model.
- Difficulty and Hash Rate are highly correlated, only one will be included in the model.

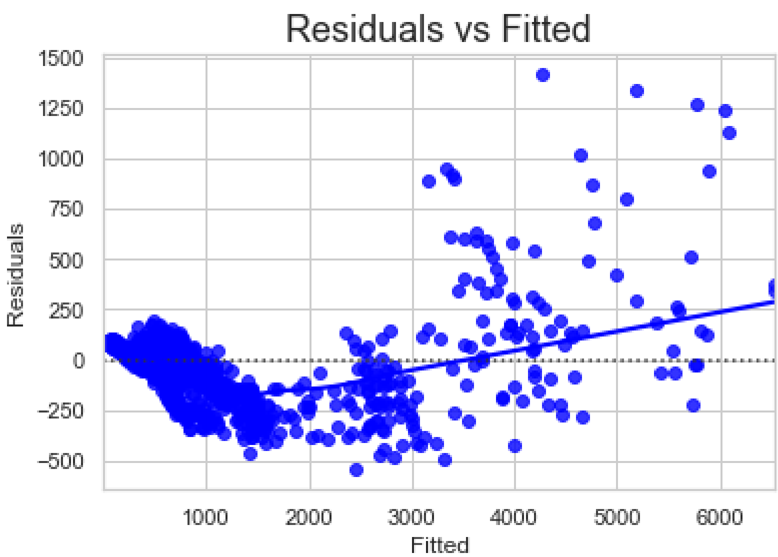
In [27]:
```
highly_correlated_variables = ["btc_difficulty",
"btc_miners_revenue", "btc_estimated_transaction_volume_usd"]
highly_correlated_variables_with_market_price =
["btc_market_price", "btc_difficulty",  "btc_miners_revenue",
"btc_estimated_transaction_volume_usd"]
# Multiple Linear Regression
# Create a new dataframe with the highly correlated variables
BTC_data3 =
BTC_data2[highly_correlated_variables_with_market_price]
# Run a multiple linear regression model using sm.OLS
BTC_data3 = sm.add_constant(BTC_data3)
fit8 = sm.OLS(BTC_data3["btc_market_price"],
```

```
BTC_data3[highly_correlated_variables]).fit()
# Print the summary
print(fit8.summary())
print(fit8)
```

```
                                 OLS Regression Results                                
=======================================================================================
Dep. Variable:       btc_market_price   R-squared (uncentered):                   0.980
Model:                            OLS   Adj. R-squared (uncentered):              0.980
Method:                 Least Squares   F-statistic:                          3.537e+04
Date:                Thu, 11 Nov 2021   Prob (F-statistic):                        0.00
Time:                        09:21:23   Log-Likelihood:                         -14135.
No. Observations:                2153   AIC:                                  2.828e+04
Df Residuals:                    2150   BIC:                                  2.829e+04
Df Model:                           3                                                  
Covariance Type:            nonrobust                                                  
=======================================================================================================
                                           coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------
btc_difficulty                        1.721e-09   3.45e-11     49.899      0.000    1.65e-09    1.79e-09
btc_miners_revenue                       0.0002    3.4e-06     68.173      0.000       0.000       0.000
btc_estimated_transaction_volume_usd  2.595e-07   3.85e-08      6.749      0.000    1.84e-07    3.35e-07
==============================================================================
Omnibus:                     1616.641   Durbin-Watson:                   0.346
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            57441.829
Skew:                           3.162   Prob(JB):                         0.00
Kurtosis:                      27.502   Cond. No.                     2.79e+05
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 2.79e+05. This might indicate that there are
strong multicollinearity or other numerical problems.
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x13547d3a0>
```

In [28]:

```
# Residuals vs fit plot
sns.residplot(x=fit8.fittedvalues, y=fit8.resid, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```



Narrow down variables some more

- Difficulty, Hash Rate and Miners revenue are highly correlated, only one will be included in the model.

In [29]:

```
highly_correlated_variables_excluding_volume =
["btc_miners_revenue", "btc_estimated_transaction_volume_usd"]
highly_correlated_variables_excluding_volume_with_market_price
=  ["btc_market_price", "btc_miners_revenue",
"btc_estimated_transaction_volume_usd"]
# Multiple Linear Regression
# Create a new dataframe with the highly correlated variables
BTC_data4 =
BTC_data2[highly_correlated_variables_excluding_volume_with_marke
```

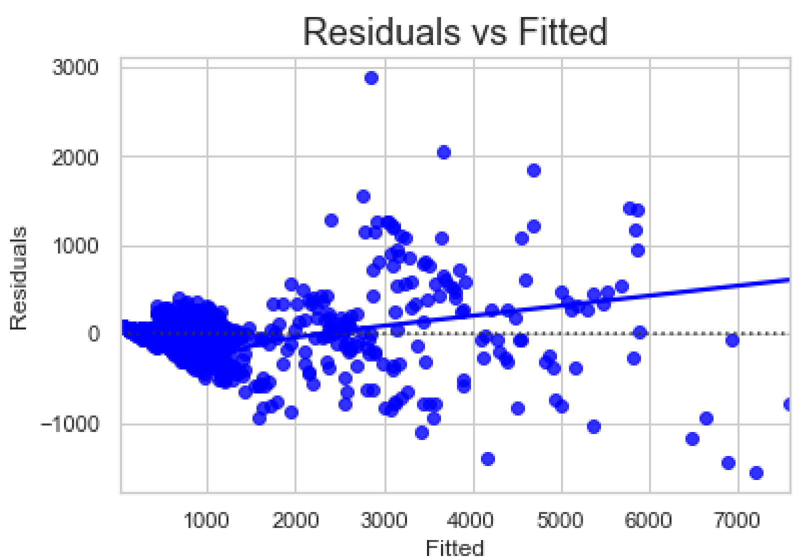
```
# Run a multiple linear regression model using sm.OLS
BTC_data4 = sm.add_constant(BTC_data4)
fit9 = sm.OLS(BTC_data4["btc_market_price"],
BTC_data4[highly_correlated_variables_excluding_volume]).fit()
# Print the summary
print(fit9.summary())
print(fit9)
```

```
                                 OLS Regression Results                                
=======================================================================================
Dep. Variable:       btc_market_price   R-squared (uncentered):                   0.957
Model:                            OLS   Adj. R-squared (uncentered):              0.957
Method:                 Least Squares   F-statistic:                          2.402e+04
Date:                Thu, 11 Nov 2021   Prob (F-statistic):                        0.00
Time:                        09:21:23   Log-Likelihood:                         -14963.
No. Observations:                2153   AIC:                                  2.993e+04
Df Residuals:                    2151   BIC:                                  2.994e+04
Df Model:                           2                                                  
Covariance Type:            nonrobust                                                  
=================================================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
btc_miners_revenue                0.0003   4.84e-06     56.597      0.000       0.000       0.000
btc_estimated_transaction_volume_usd  1.491e-06   4.33e-08     34.423      0.000    1.41e-06    1.58e-06
==============================================================================
Omnibus:                     1510.442   Durbin-Watson:                   0.588
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            58848.695
Skew:                           2.811   Prob(JB):                         0.00
Kurtosis:                      27.988   Cond. No.                         275.
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x1350b8
5b0>
```

In [30]:

```
# Residuals vs fit plot
sns.residplot(x=fit9.fittedvalues, y=fit9.resid, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```

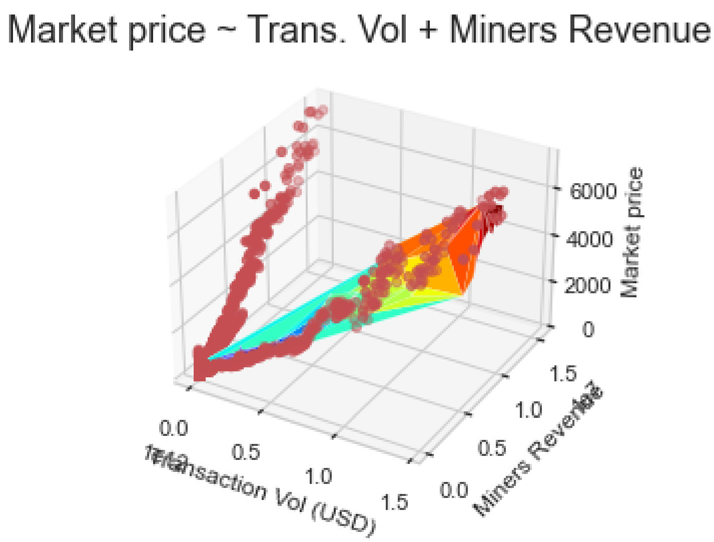


Despite some heteroscedasticity this is the best model so far

In [31]:

```
# Draw a  3D plot of the regression line
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
# Plot the surface.
ax.scatter(BTC_data2["btc_difficulty"],
BTC_data2["btc_hash_rate"], BTC_data2["btc_market_price"],
c='r', marker='o')
ax.plot_trisurf(BTC_data2["btc_difficulty"],
BTC_data2["btc_hash_rate"], fit8.fittedvalues, cmap='jet',
```

```
linewidth=0.1)
ax.scatter(BTC_data2["btc_estimated_transaction_volume_usd"],
BTC_data2["btc_miners_revenue"], BTC_data2["btc_market_price"],
c='r', marker='o')
plt.title('Market price ~ Trans. Vol + Miners Revenue')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
ax.set_xlabel('Transaction Vol (USD)')
ax.set_ylabel('Miners Revenue')
ax.set_zlabel('Market price')
ax.set_zlim(0, BTC_data2["btc_market_price"].max())
plt.show()
```



Polynomial Multilinear Regression

Linear Model: Market Price ~ Miners Revenue Squared + Count of Transactions Squared

In [32]:
```
# Create a dataframe with the highly correlated variables
BTC_data5 = BTC_data2[[ "btc_market_price",
"btc_estimated_transaction_volume_usd", "btc_miners_revenue",
"Date"]]
# Convert BTC_data5["btc_miners_revenue"] to a degree-2
polynomial
BTC_data5 = BTC_data5.loc[:]
BTC_data5["btc_miners_revenue_squared"] = BTC_data5.loc[:,
```

```
"btc_miners_revenue"].apply(lambda x: np.power(x, 2))
# Convert BTC_data5["btc_estimated_transaction_volume_usd"] to a
degree-2 polynomial
BTC_data5["btc_estimated_transaction_volume_usd_squared"] =
BTC_data5.loc[:,
"btc_estimated_transaction_volume_usd"].apply(lambda x:
np.power(x, 2))
# Run a multiple linear regression model using sm.OLS
BTC_data5 = sm.add_constant(BTC_data5)
fit10 = sm.OLS(BTC_data5["btc_market_price"],
BTC_data5[["btc_estimated_transaction_volume_usd",
"btc_miners_revenue"]]).fit()
# Print the summary
print(fit10.summary())
print(fit10)
```

```
                                 OLS Regression Results
=================================================================================
=======
Dep. Variable:       btc_market_price   R-squared (uncentered):
0.957
Model:                            OLS   Adj. R-squared (uncentered):
0.957
Method:                 Least Squares   F-statistic:                           2.
402e+04
Date:                Thu, 11 Nov 2021   Prob (F-statistic):
0.00
Time:                        09:21:24   Log-Likelihood:
-14963.
No. Observations:                2153   AIC:                                   2.
993e+04
Df Residuals:                    2151   BIC:                                   2.
994e+04
Df Model:                           2
Covariance Type:            nonrobust
=================================================================================
========================
                                          coef    std err          t      P>|t|
[0.025      0.975]
---------------------------------------------------------------------------------
------------------------
btc_estimated_transaction_volume_usd  1.491e-06   4.33e-08     34.423      0.000
1.41e-06    1.58e-06
btc_miners_revenue                       0.0003   4.84e-06     56.597      0.000
0.000       0.000
===============================================================================
Omnibus:                     1510.442   Durbin-Watson:                      0.588
```

```
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           58848.695
Skew:                           2.811   Prob(JB):                        0.00
Kurtosis:                      27.988   Cond. No.                        275.
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not conta
in a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
<statsmodels.regression.linear_model.RegressionResultsWrapper object at 0x137798
e80>
```

In [33]:

```
# Residuals vs fit plot
sns.residplot(x=fit10.fittedvalues, y=fit10.resid, lowess=True,
color='blue')
plt.title('Residuals vs Fitted')
plt.xlabel('Fitted')
plt.ylabel('Residuals')
plt.rc('font', size=14)
plt.rc('figure', titlesize=18)
plt.rc('axes', labelsize=12)
plt.rc('axes', titlesize=18)
plt.show()
```