

Excel Composite Index Construction

20 Feb 2022

## Growth Composite Index – U.S. Market

**Interview Candidate:**

**Andy Chan**  
andyinter1@gmail.com  
+852 9603 3105

## Growth Composite Index Construction

### Purpose

Form a composite index to determine the economic/GDP growth trend of U.S.

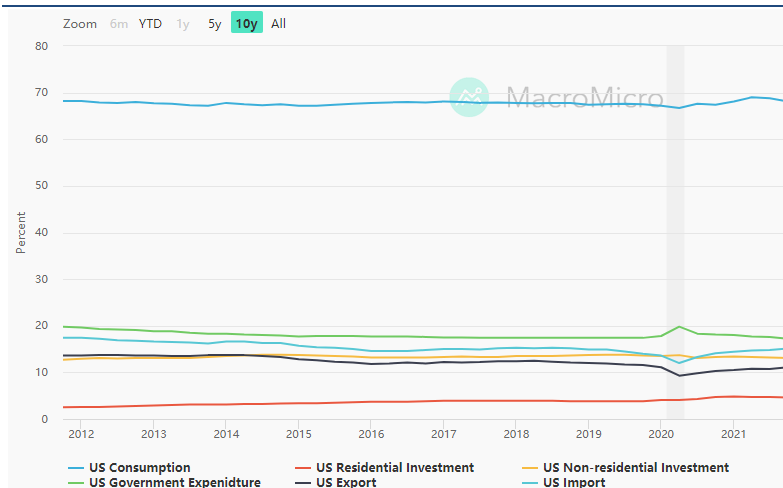
### Data Selection

We first pre-screen and identify potential component candidates to be included in the composite index from the economic indicator pools in website such as Trading Economics.

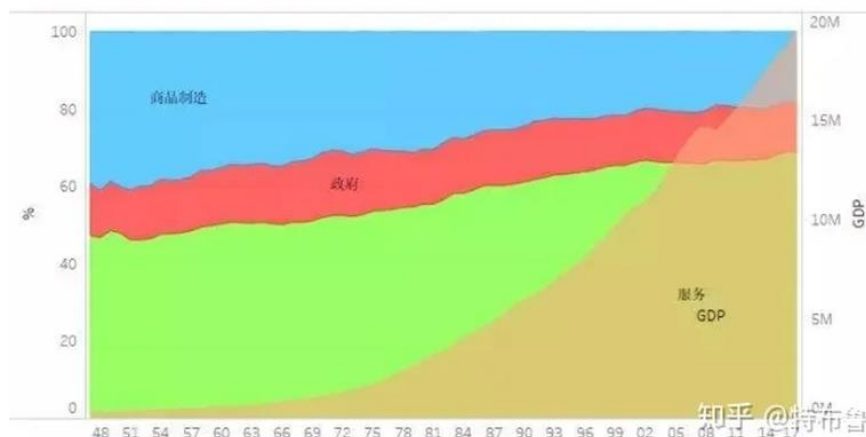
Indicators will be searched based on 4 dimensions: 1) consumption, 2) business, 3) government, and 4) capital market. To further narrow the searching direction of indicators, consumption dimension is divided into “ability to spend” and “willingness to spend”, while business dimension is divided into “service sector” and “manufacturing sector”.

Number of indicators selected for each dimension will be based on the relative importance of the dimension. For example, given that almost 70% of the U.S. GDP is contributed by consumption according to the GDP structure of U.S. (fig. 1), number of indicators picked for consumption dimension will be the highest.

**Figure 1 Economic structure of U.S.**



美国产业发展



**Index indicator components filtered**

Dimension	Sub-dimension	Indicator
Consumption	• Ability to spend	<ul style="list-style-type: none"> <li>Initial Jobless Claims</li> <li>U.S. Avg. Weekly Working Hours</li> <li>U.S. Avg. Hourly Earnings</li> </ul>
	• Willingness to spend	<ul style="list-style-type: none"> <li>Uni of Michigan: Consumer Sentiment Index</li> </ul>
Business	• Service conditions	<ul style="list-style-type: none"> <li>U.S. ISM Non-Manufacturing New Orders</li> </ul>
	• Manufacturing conditions	<ul style="list-style-type: none"> <li>Manufacturers' New Orders Total Manufacturing</li> </ul>
Gov. Policy	• Interest rate	<ul style="list-style-type: none"> <li>10yrs/3mths interest spread</li> </ul>
Capital Mkt	• Stock market	<ul style="list-style-type: none"> <li>S&amp;P 500</li> </ul>

Once potential indicator candidates have been filtered out, each data series will be examined mainly based on 1) economic relevance, 2) timeliness, 3) availability, 4) reliability. For example, data series with only quarterly data will be excluded.

**Data series evaluation**

<b>Initial Jobless Claims</b>	<b>Economic relevance</b>	<ul style="list-style-type: none"> <li>➤ Higher income to spend in terms of whole population if fewer citizens unemployed</li> <li>➤ Higher hiring needs reflect stronger business demand</li> </ul>
	<b>Timeliness</b>	<ul style="list-style-type: none"> <li>➤ Every Thursday, publish data of previous week</li> <li>➤ Max time gap between latest publication and current day: 1 week</li> </ul>
	<b>Availability</b>	<ul style="list-style-type: none"> <li>➤ Weekly data</li> <li>➤ With 3 years data records from 2019 to 2022</li> </ul>
	<b>Reliability</b>	<ul style="list-style-type: none"> <li>➤ From U.S. Department of Labor</li> </ul>
<b>U.S. Avg. Weekly Working Hrs of All Employees, Total Private</b>	<b>Economic relevance</b>	<ul style="list-style-type: none"> <li>➤ Higher working hours, higher income to spend</li> <li>➤ Higher working hours reflect stronger business demand</li> </ul>
	<b>Timeliness</b>	<ul style="list-style-type: none"> <li>➤ Start of each month, publish data of previous month</li> <li>➤ Max time gap between latest publication and current day: 1 month</li> </ul>
	<b>Availability</b>	<ul style="list-style-type: none"> <li>➤ Monthly data</li> <li>➤ With 3 years data records from 2019 to 2022</li> </ul>
	<b>Reliability</b>	<ul style="list-style-type: none"> <li>➤ From U.S. Bureau of Labor Statistics</li> </ul>
<b>U.S. Avg. Hourly Earnings of All Employees, Total Private</b>	<b>Economic relevance</b>	<ul style="list-style-type: none"> <li>➤ Higher hourly earnings, higher income to spend</li> <li>➤ Rising wage to keep employees indicates strong hiring demand, which reflects stronger business demand</li> </ul>
	<b>Timeliness</b>	<ul style="list-style-type: none"> <li>➤ Start of each month, publish data of previous month</li> <li>➤ Max time gap between latest publication and current day: 1 month</li> </ul>
	<b>Availability</b>	<ul style="list-style-type: none"> <li>➤ Monthly data</li> <li>➤ With 3 years data records from 2019 to 2022</li> </ul>
	<b>Reliability</b>	<ul style="list-style-type: none"> <li>➤ From U.S. Bureau of Labor Statistics</li> </ul>

<b>Uni of Michigan: Consumer Sentiment Index</b>	<b>Economic relevance</b>	➤ Being confident about future economy increases current spending
	<b>Timeliness</b>	➤ Prelim result: mid of each month ➤ Final result: end of each month ➤ Max time gap between latest publication and current day: ~ 15 days
	<b>Availability</b>	➤ Monthly data ➤ With 3 years data records from 2019 to 2022
	<b>Reliability</b>	➤ From University of Michigan
<b>U.S. ISM Non- Manufacturing New Orders</b>	<b>Economic relevance</b>	➤ More orders supported by higher economic vitality and demand
	<b>Timeliness</b>	➤ Start of each month, publish data of previous month ➤ Max time gap between latest publication and current day: 1 month
	<b>Availability</b>	➤ Monthly data ➤ With 3 years data records from 2019 to 2022
	<b>Reliability</b>	➤ From Institute of Supply Management
<b>Manufacturers' New Orders Total Manufacturing</b>	<b>Economic relevance</b>	➤ More orders supported by higher economic vitality and demand
	<b>Timeliness</b>	➤ No Feb and Jan data for 2022
	<b>Availability</b>	➤ Monthly data ➤ With data records from 2019 to 2021
	<b>Reliability</b>	➤ From U.S. Census Bureau
<b>10yrs/3m interest spread</b>	<b>Economic relevance</b>	➤ Decreasing short term rate stimulates investment and consumption ➤ Increasing long term rate indicates confidence in economic rebound
	<b>Timeliness</b>	➤ Real time data
	<b>Availability</b>	➤ Daily Data ➤ With 3 years data records from 2019 to 2022
	<b>Reliability</b>	➤ From U.S. Treasury Department
<b>S&amp;P 500</b>	<b>Economic relevance</b>	➤ Investors switch from bonds to equity if they feel confidence in economic growth
	<b>Timeliness</b>	➤ Real time data
	<b>Availability</b>	➤ Daily Data ➤ With 3 years data records from 2019 to 2022
	<b>Reliability</b>	➤ From S&P Dow Jones Indices LLC

After further evaluation:

- Manufacturers' New Orders Total Manufacturing is replaced by U.S. ISM Manufacturing New Orders published by Institute of Supply Management due to lack of data availability and timeliness.
- U.S. Avg. Hourly Earnings is removed due to relatively low correlation between it and GDP growth.
- U.S. Avg. Weekly Working Hours is replaced by Weekly Hours Worked: Manufacturing for the United States published by OECD due to higher correlation between it and GDP growth.

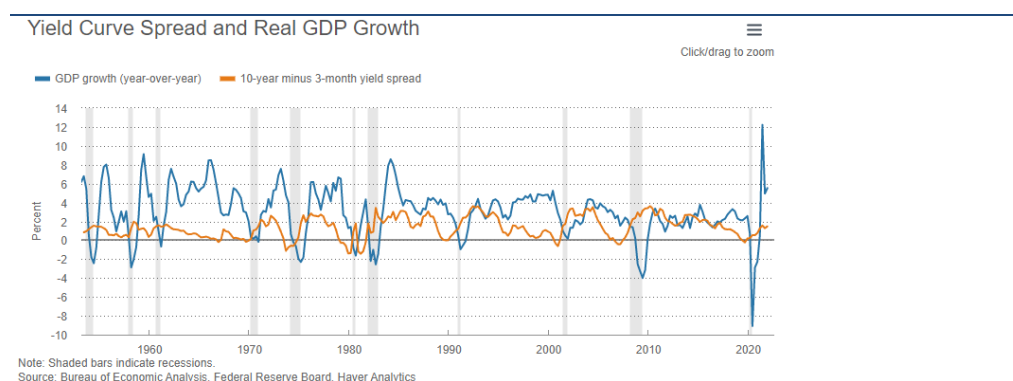
**Chosen index components after evaluation**

Dimension	Sub-dimension	Indicator
Consumption	• Ability to spend	• Initial Jobless Claims • Weekly Hrs Worked Manufacturing for U.S.
	• Willingness to spend	• Uni of Michigan: Consumer Sentiment Index
Business	• Service conditions	• U.S. ISM Non-Manufacturing New Orders
	• Manufacturing conditions	• U.S. ISM Manufacturing New Orders Index
Gov. Policy	• Interest rate	• 10yrs/3m interest spread
Capital Mkt	• Stock market	• S&P 500

**Data Treatment**

MoM symmetric percent changes of various data series are the major inputs to form the composite index. Other than interest rate spread series, MoM symmetric percent changes will be calculated for all data series.

For interest rate series, based on Figure 2, not the growth rate of interest spread but the spread itself moves along with GDP growth except during recession time. Therefore, the level of spread, but not its MoM growth, will be used to form composite index.

**Figure 2 Yield spread and GDP growth relationship****Imputation of missing data**

Three index components, namely Weekly Hours Worked Manufacturing for the United States, U.S. ISM Manufacturing New Orders, and U.S. ISM Non-Manufacturing New Orders, miss their Feb 2021 data (fig. 3).

Regarding to dealing with missing data of indicator because of late publication, instead of ignoring the missing data or eliminating Feb data from all data series, data imputation method is applied. We apply mean imputation method which replace the empty data cell with average growth rate of previous 6 months. There is no other missing data for all data series from Jan 2019 to Jan 2021.

Figure 3 Example of missing data in Weekly Hours Worked data series

Weekly Hours Worked Manufacturing for the United States (FRED)							
Mean	-0.0600						
Median	0.0000						
MAD	0.48						
SD	1.3498						
Min	-7.2773						
Max	2.5707						
I-S weight	0.2638						
Date	Data	Growth MoM	Z-score	Max Min	Abs Diff	Robust Z-score	
1/2/2022	41.2	-0.121	-0.68	0.20	0.48	-1.93	
1/1/2022	41.2	-0.242	-0.85	0.18	0.60	-2.42	

It should be noted that the missing pattern is not “missing completely at random” as the value is correlated with growth data in previous period and linear regression method should be a better method. However, as only 1 sample value is missing in each data series, mean imputation is used for simplification purpose.

#### Data Normalization

Considering that symmetric percent changes MoM of data series are used to form composite index, loosely speaking, the data series are the same variable on equal ground. The only exception is interest rate spread, which is also a relative variable (yield difference between 10yrs and 3mths treasury) presented in percentage term. Therefore, symmetric percent changes of data series will be used directly without additional transformation. Please refer to “Another possible method to form composite index” section for further information of normalization method application.

#### Outlier detection

Regarding to dealing with outliers, we first look at mean and median of various data sets. If significant inconsistency is identified (fig. 4), we will filter out any data record with a robust z-score  $>3$  /  $<-3$  (fig. 5).

Given that all the outliers (mostly due to COVID outbreak) are legitimate value rather than data error, outliers will not be trimmed or winsorized so that their exceptional performance will be recognized in the index.

Figure 4 Outlier detection – mean vs median

University of Michigan: Consumer Sentiment Index (FRED)	
Mean	-1.2222
Median	0.5981
MAD	3.33
SD	6.5075
Min	-21.5040
Max	10.0186
I-S weight	0.0547

Figure 5 Outlier detection - Robust Z-score

Date	Data	Rebased	Growth MoM	Z-score	Max Min	Abs Diff	Robust Z-score
1/5/2020	72.3	73.55	0.694	0.29	0.70	0.10	0.02
1/4/2020	71.8	73.04	-21.504	-3.12	0.00	22.10	-4.48
1/3/2020	89.1	90.64	-12.520	-1.74	0.29	13.12	-2.66

## Weighting

We apply Iyengar-Sudarshan method to calculate weights to ensure that large variation in any one of the indicators would not impose a dominant effect on the final composite index. The lower the standard deviation a data series has, the higher the weight it is assigned.

**Figure 6 Standard deviation difference between data series**

Initial Jobless Claim (FRED)		Weekly Hours Worked Manufacturing for the United States (FRED)	
Mean	-3.5093	Mean	-0.0600
Median	-3.0019	Median	0.0000
MAD	8.90	MAD	0.48
SD	36.6171	SD	1.3498
Min	-75.3386	Min	-7.2773
Max	186.6688	Max	2.5707
I-S weight	0.0097	I-S weight	0.2638

The lower the SD, the higher the weight

## Data Aggregation

Every symmetric percent change MoM data in each data series will multiply its respective Iyengar-Sudarshan weight to calculate its marginal contribution. For each month in the sample period, corresponding marginal contributions are summed and become the composite monthly growth of the growth composite index.

Every composite monthly growth will then be adjusted to match the growth trend of U.S. GDP. The average MoM growth of Coincident Economic Activity Index for the United States is used as a proxy of U.S. GDP growth. The difference between the MoM growth of Coincident Economic Activity Index and the growth composite index will be added into every composite monthly growth calculated.

Finally, composite index is formed based on adjusted composite monthly growth, with Jan 2019 as the base month.

**Figure 7 Growth Composite Index formed**

Index using Iyengar-Sudarshan weightage method		
Composite Monthly % Chg	Adjusted Composite Monthly % Ch	Growth Composite Index
-0.14	-0.44	105.23
0.69	0.39	105.69
0.80	0.50	105.28
1.19	0.89	104.75
0.35	0.05	103.83
1.27	0.97	103.78
-0.04	-0.34	102.78
0.80	0.50	103.13
1.13	0.83	102.62
1.23	0.93	101.77
0.94	0.64	100.82
2.61	2.31	100.18
0.07	-0.23	97.89
0.49	0.19	98.12
1.40	1.10	97.93
-0.14	-0.44	96.86
0.56	0.26	97.28
1.35	1.05	97.03
0.97	0.67	96.02
1.10	0.80	95.38
4.51	4.21	94.61
3.58	3.28	90.71
-8.28	-8.58	87.78
-2.11	-2.41	95.65
0.44	0.13	97.98
0.94	0.64	97.85
0.05	-0.25	97.23
0.41	0.11	97.47
0.21	-0.09	97.37
-0.45	-0.75	97.45
-0.57	-0.87	98.18
0.26	-0.04	99.04
-0.81	-1.11	99.08
0.29	-0.01	100.18
-0.20	-0.50	100.19
0.53	0.23	100.69
0.76	0.46	100.46
-0.86	-1.16	100.00

## Comparison

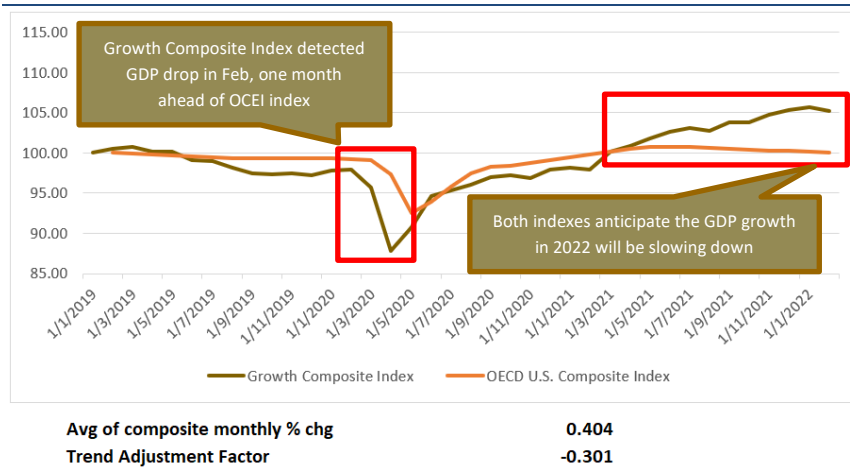
OCEI composite index is picked as a benchmark for comparison.

Growth Composite Index does not only detect a significant drop but also a rapid rebound in anticipated GDP one month ahead of OCEI composite index (fig. 8). The drop signal generated by Growth Composite Index would be sent out in Feb 20, one month before COVID-19 international outbreak in Mar 20.

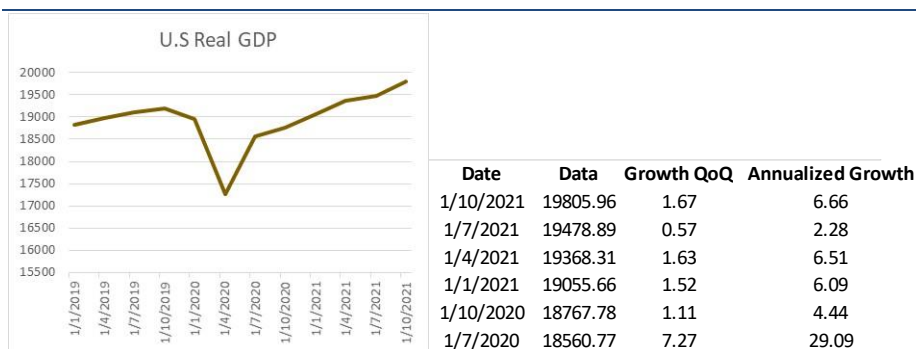
Apart from that, based on real GDP data from 2009 to 2021, the mean of annualized growth rate of real GDP is 2.04%. Therefore, the GDP annualized growth rates since 2Q20 (fig. 9) are above the long-term trend. However, the OCEI index stayed at around 100 since late 2020, anticipating that the U.S GDP only grew at its trend rate in 2021, which is not accurate. In contrast, the Growth Composite Index keep rising since late 20.

On the other hand, the MoM decreases in both indexes (fig. 10) imply that the anticipated real GDP growth in 2022 would decelerate.

**Figure 8 Growth composite index with Iyengar-Sudarshan weightage method**



**Figure 9 U.S. Real GDP (quarterly data)**



**Figure 10 GDP growth is anticipated to be slowing down**

Index using Iyengar-Sudarshan weightage	OCEI MoM Growth
Adjusted Composite Monthly % Chg	
-0.44	-0.066
0.39	-0.068
0.50	-0.076
0.89	-0.096
0.05	-0.117
0.97	-0.130



## Another possible method to form composite index

Alternatively, the Growth Composite Index can be formed with Z-score data.

### Data Normalization

MoM symmetric percent changes of data series will be normalized. As outliers exist, normalization methods such as min max technique and distance to reference are not appropriate. Given that those outliers are legitimate value, Z-score normalization will be used as the impact of extreme values are not as influential as in other methods.

### Weighting

Equal weighting is used for this index. As a result, composite normalized MoM change is calculated as an average of MoM symmetric percent changes Z-score in the corresponding month. Composite standard deviation and composite mean is respectively calculated as an average of standard deviation and mean of each data series.

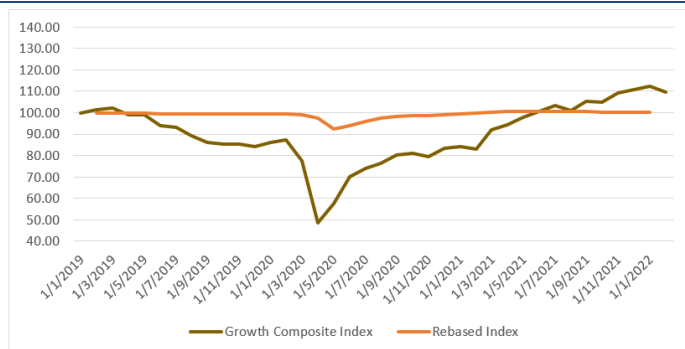
### Data Aggregation

As indicator data are not strictly positive (negative MoM changes exist), arithmetic aggregation rather than geometric aggregation is more appropriate. Z-score of MoM growth will be summed up and become composite normalized monthly growth.

With composite standard deviation and composite mean, we can reversely calculate monthly growths of the composite index from its normalized monthly growths. The Growth Composite Index is then formed with the new monthly growths.

The difference between the MoM growth of Coincident Economic Activity Index and the growth composite index will be added into every composite monthly growth calculated.

**Figure 11 Growth composite index with z-score normalization**



Composite Amplitude	11.200
Composite Trend	-0.376
Avg of composite monthly % chg	-1.33
Trend Adjustment Factor	1.434