Quantitative Survey of Cybersecurity in Periods of Economic Contraction

Cheaheon “Ian” Lim

Version: 5.24.2020

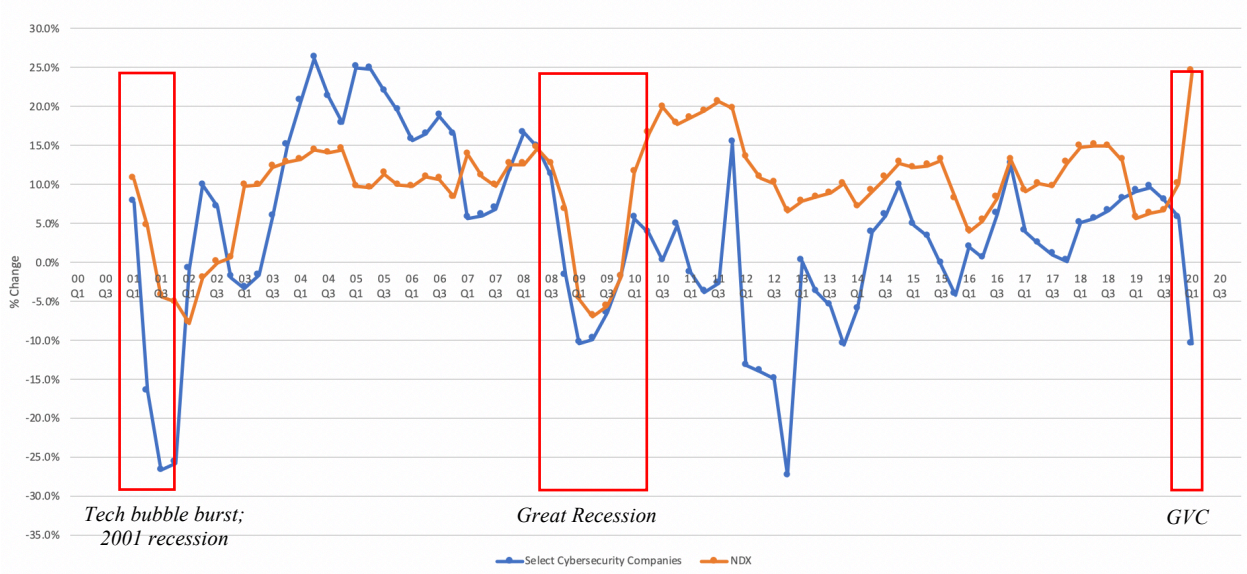
Cheaheon "Ian" Lim
ilim@college.harvard.edu

I. Benchmarking Industry Performance

Framework

To gain a preliminary understanding of the cybersecurity industry during periods of economic contraction, the YoY change in the average quarterly revenue of 35 publicly traded cybersecurity firms (partially sourced from the CIBR ETF; complete list in appendix) were compared to that of companies in the Nasdaq-100 index.

Figure 1: YoY change quarterly revenue for cybersecurity firms versus NDX



** A handful of firms have yet to report Q1 2020 earnings*
Source(s): S&P Global Market Intelligence

Divergence in Performance: Tech Bubble versus Great Recession

As can be seen in Figure 1, the average revenue of the selected cybersecurity firms moved in line with that of Nasdaq-100 companies during the Great Recession, while the decrease in annual revenue was more pronounced than the index benchmark in the quarters that followed the tech bubble burst. This phenomenon points to the fact that national cybersecurity spend decreases not only in periods of economic contraction, but also in bear markets—during which the management team of technology companies tend to signal investors by widening margins (which can be attained by cutting cybersecurity spend, among other expenses).

The Great Virus Crisis (GVC)

Government-mandated shelter-in-place measures began in late March, and as a result, the GVC did not have a significant impact on Q1 revenues for most Nasdaq-100 companies. However, even before COVID-19 started noticeably impacting firm-level operations in the last few weeks of the first quarter, many companies started implementing cost-cutting measures to bolster cash balances and brace for the GVC. Such actions most likely involved reducing spend on cybersecurity-related services, as is indicated by the noticeable ~10% YoY decline in the

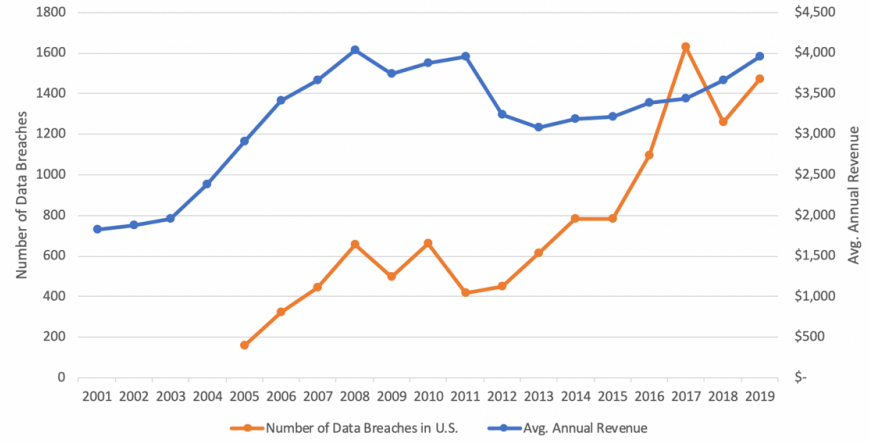
quarterly revenue of cybersecurity firms in Figure 1 (note that a number of prominent names in the analyzed cybersecurity basket—including BAH, OKTA, VMW, and ZS—have yet to report earnings as of 5/24/2020).

II. Modeling Annual Number of Data Breaches

Framework

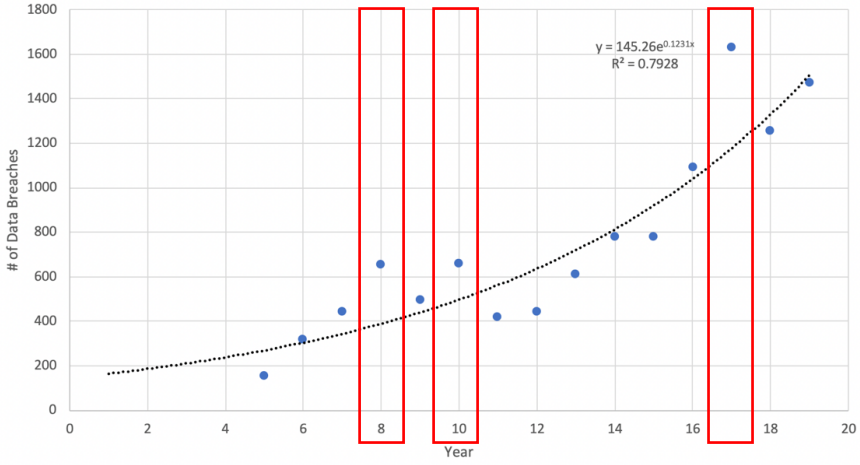
An exponential regression was performed on the annual number of data breaches in the U.S. with the underlying assumption that the number of cyberattacks experiences baseline organic growth YoY regardless of external factors—this makes intuitive sense as web-based technology increasingly becomes an integral part of most firms' operations. The present section seeks to determine the correlation between national spend on cybersecurity (revenue of selected cybersecurity firms used as a proxy variable) and deviations from model-driven expectations for the annual number of data breaches.

Figure 2: Selected cybersecurity firms' avg. yearly revenue and number of data breaches in the U.S. over time



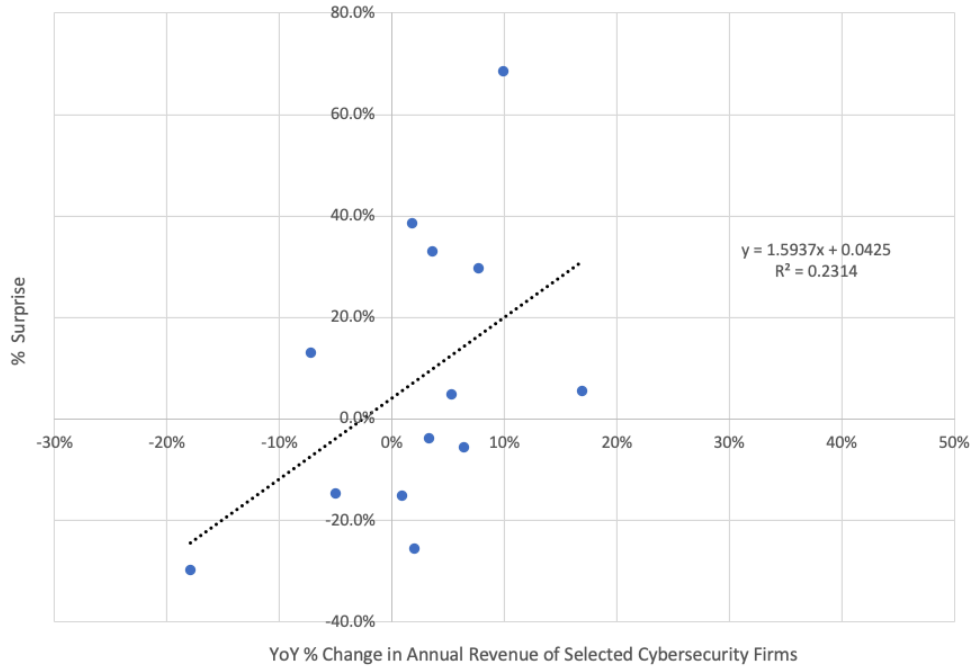
Source(s): Identity Theft Resource Center, S&P Global Market Intelligence

Figure 3: Number of data breaches in U.S. over time (noted as years since 2000)



Source(s): Identity Theft Resource Center

Figure 4: % Difference in expected and actual number of data breaches as a function of YoY % change in avg. annual revenue of select cybersecurity firms



Source(s): Identity Theft Resource Center, S&P Global Market Intelligence

Exponential Model for Data Breaches

As can be seen in Figure 3, the annual number of data breaches in the U.S. closely follows an exponential growth model, where *x* denotes the number of years since 2000:

$$f_{expected}(x) = 145.25 * e^{0.123x}$$

In the years 2008, 2010, and 2017 (highlighted by red boxes), the actual number of data breaches noticeably surpassed model expectations for the year, due to a number of potential extraneous factors further elaborated on in Section III. This "deviation" from model-expectations were quantified using the *% surprise* variable, which is used in Figure 4 and subsequent models:

$$deviation = actual\ \#\ of\ data\ breaches - f_{expected}(x)$$

$$\%\ surprise = \frac{deviation}{f_{expected}(x)}$$

Figure 4 demonstrates that *% surprise* is positively correlated with the YoY % change in revenue of selected cybersecurity firms, demonstrating that an increase in cybersecurity-related spending (i.e. increase in revenue of cybersecurity firms) may be a reactionary consequence of a higher-than-expected number of data breaches (i.e. higher *% surprise*) in a given year.

Cheaheon "Ian" Lim
ilim@college.harvard.edu

III. Correlation between Unemployment and Cybersecurity

Framework

Section II demonstrated that the YoY % change in revenue of cybersecurity firms could not be used to model for *% surprise*, as the latter variable seemed to impact the former, rather than the other way around. In this section, the annual unemployment rate is used as a proxy variable to determine the potential existence of a causational relationship between the macroeconomic climate and *% surprise.*

Unemployment rate (and as later described, its YoY nominal change) is a more appropriate variable for performing this analysis, especially when forecasting GVC-induced implications for 2020, than other measures of economic activity such as real GDP growth. This is because the unemployment rate reflects not only the general macro climate (which pertains to the cybersecurity services industry as described in Sections I and II), but also the conditions for in-house employees that specialize in cybersecurity. In light of budget cuts and hiring freezes that many managers are facing due to the GVC, it seems reasonable to assume that the GVC's effect on cybersecurity is going to be twofold: (1) decrease in cybersecurity services industry revenue, and (2) decrease in in-house employment of cybersecurity personnel.

Figure 5: % Surprise versus unemployment rate over time

Source(s): Identity Theft Resource Center, St. Louis Fed

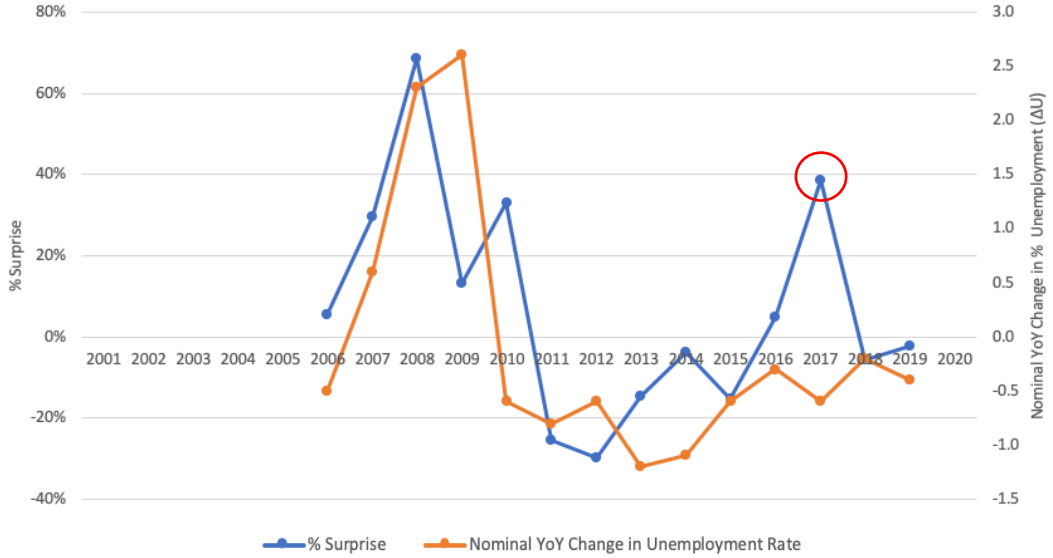
As can be seen in Figure 5, the YoY (nominal) change in annual unemployment seemed to roughly coincide with the *% surprise* variable, with *% surprise* reaching near-0 levels as unemployment rates continued to decrease over time (the exception is 2009 and 2017).

Cheaheon "Ian" Lim
ilim@college.harvard.edu

Accounting for this observation, Figure 6 tracks the *% surprise* and YoY nominal change in unemployment rate over time. However, instead of using annualized unemployment rates, the YoY nominal change in unemployment rate " ΔU " was defined in the following manner:

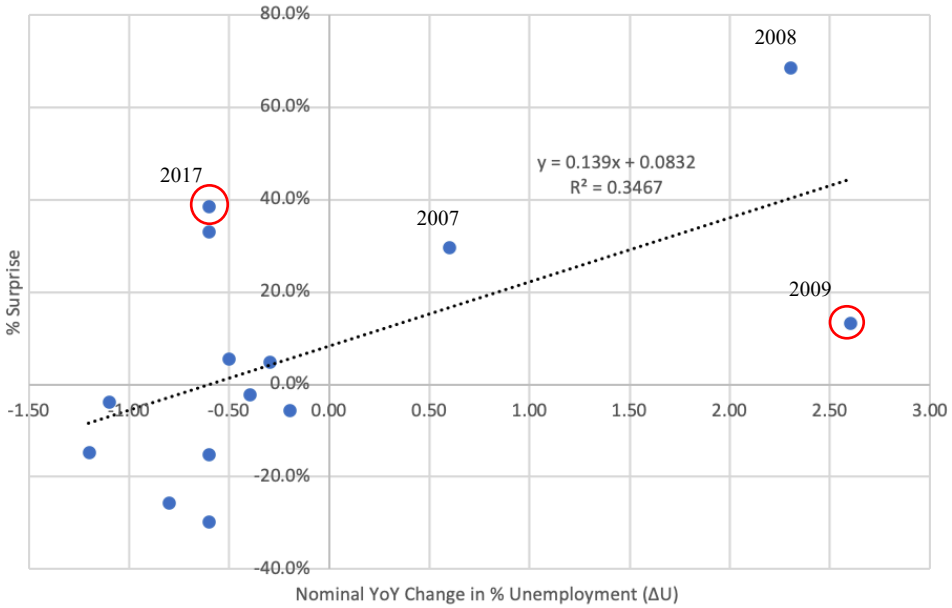
$$\Delta U_{year} = U_{final} - U_{initial}$$
$$U_{final} = \%\ unemployment\ in\ Dec.(current\ year)$$
$$U_{initial} = \%\ unemployment\ in\ Dec.(previous\ year)$$

Figure 6: % Surprise versus yearly change (nominal) in unemployment rate



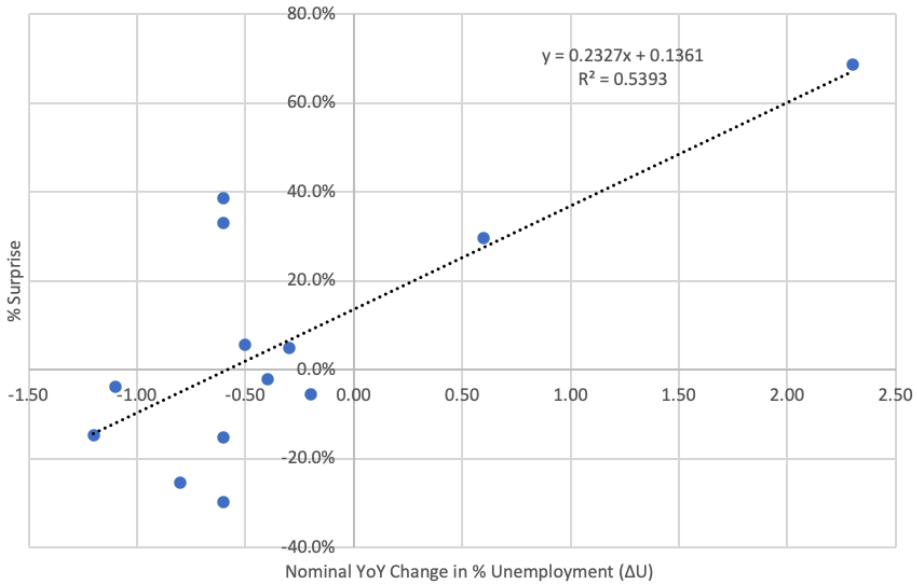
Source(s): Identity Theft Resource Center, St. Louis Fed

Figure 7: % Surprise as a function of YoY nominal change in unemployment rate



Source(s): Identity Theft Resource Center, St. Louis Fed

Figure 8: % Surprise as a function of YoY nominal change in unemployment rate (2009 datapoint deleted)



Source(s): Identity Theft Resource Center, St. Louis Fed

The 2017 Datapoint

In 2017 (refer to Figure 6), although ΔU was at sub-zero levels, the % *surprise* was relatively high, which equates to the fact that the number of data breaches was much higher than model expectations (from Section II) in 2017 despite a YoY decrease in unemployment rates. This is in line with the observation made in Section II on Figure 3, where the number of data breaches in 2017 were surprisingly higher even when compared to subsequent years. One potential explanation for this anomaly is the untethered ransomware outbreaks (WannaCry, NotPetya, and Bad Rabbit) that occurred in 2017. None of these entities were originally designed to cause a global outbreak, but the ransomware strains ended up spreading far beyond the creators' intentions due to an exploit that they did not fully understand at the time.[1]

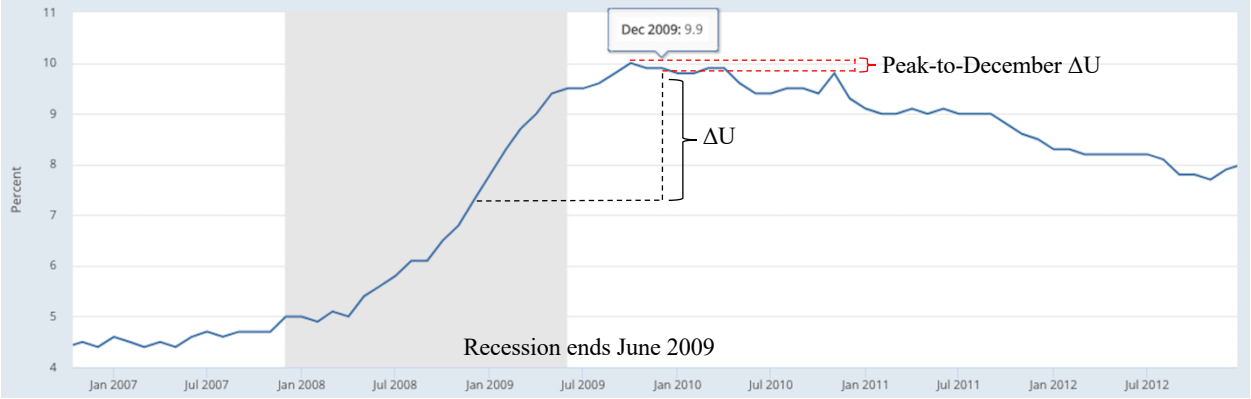
The 2009 Datapoint

As can be seen in Figures 7 and 8, there exists a noticeable positive linear correlation between YoY nominal change in % unemployment and *% surprise*. The strength of the relation is significantly degraded by the 2009 datapoint, and this can largely be attributed to the method the model uses to calculate ΔU. For all other datapoints, ΔU aptly captures the momentum of macroeconomic conditions throughout the year because the inflection (or "peak") point observed in 2009 does not exist for other years. In 2009, however, because the Great Recession ended mid-year, the unemployment rate peaks in October (10.0%) and declines 0.1% by December (9.9%). When using the ΔU method, the recovery that followed after the recession ended in June is not fully captured for 2009 due to the fact that unemployment rates were still in the process of rapidly climbing at year-end 2008 (refer to Figure 9).

[1] https://www.zdnet.com/article/a-decade-of-hacking-the-most-notable-cyber-security-events-of-the-2010s/

Cheaheon "Ian" Lim
ilim@college.harvard.edu

Figure 9: Unemployment rate over time — demonstration of "peak-to-December" versus ΔU calculation



Source(s): St. Louis Fed

Linear Model for Projecting % Surprise as a Function of ΔU

In terms of making projections for 2020 in light of the GVC, it seems appropriate to leave the 2009 datapoint as is because prevailing consensus is that the economy will "bottom-out" in Q2 2020.[2] Then, the linear model for *% surprise* as a function of ΔU (from Figure 7) is as follows:

$$\%\ surprise = 0.139\ (\Delta U) + 0.08$$

Examining the general distribution of datapoints in Figure 7 follows general intuition that crime rates go up in times of economic distress: in years of economic growth when the unemployment rate is decreasing (i.e. ΔU is negative), the *% surprise* tends to be negative or close to zero, which means that the actual number of data breaches is less than or equal to model-driven expectations (from Section II) for the year. On the converse, when unemployment rates are climbing and ΔU is positive (especially in 2007 and 2008), *% surprise* tended to be higher, with a more-than-expected number of data breaches happening in a given year.

Prior to continuing to Section IV (where various projections for 2020 are presented), it is important to note the limitations of the data that was used to generate the models. Cybercrime is a relatively novel phenomenon that has only been a major concern in recent decades, and the only datapoints we have with positive ΔU come from the Great Recession (refer to Figure 7). With unemployment rates skyrocketing as a result of the GVC, the projected ΔU for 2020 is far greater than anything we have seen in 2008 and 2009 (refer to Figures 10 and 11). As popular media has been reporting for many weeks, the situation is indeed unprecedented and there is no way to know if the models derived in the previous sections will hold for 2020 and the GVC. The model is also sensitive to extraneous factors such as a second wave of infections putting an upward pressure on unemployment rates in the Fall.

[2] https://www.cbo.gov/publication/56335

Cheaheon "Ian" Lim
ilim@college.harvard.edu

IV. Scenario Testing for 2020

Framework

Using the exponential model derived in Section II and linear model derived in Section III, a series of scenarios can be projected out for 2020 based on economists' expectations for unemployment rates. In the present moment, the GVC has led to unemployment rates reaching a record-high of 14.7% in April—but it is important to recognize that 88% of layoffs were reported as temporary, with expectations that positions will reopen once the nation returns to normalcy. Pulling from a range of projections for the annual unemployment rate for 2020, this section provides high/base/low case scenarios for the expected number of data breaches for the present year, assuming that the models hold for the unprecedented impact of the GVC.

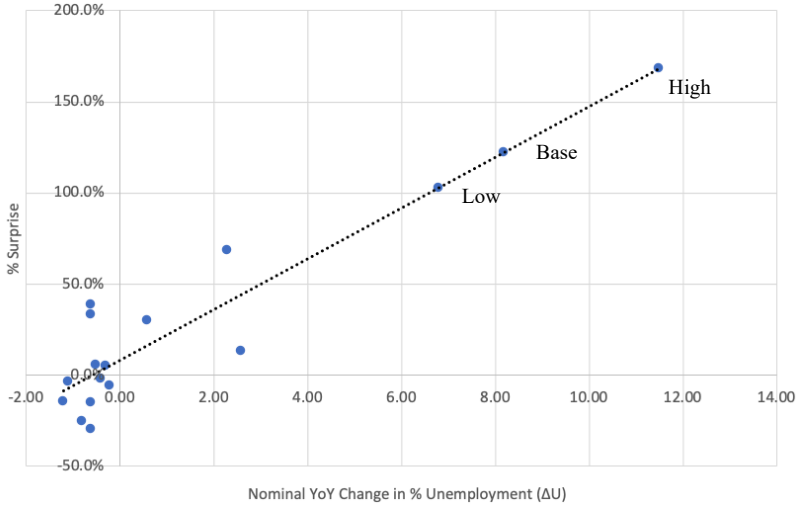
Scenario Assumptions

As of April 24th, the Congressional Budget Office (CBO) projects the unemployment rate to be 14.0% in Q2, 16.0% in Q3, and 11.7% in Q4. Using the 11.7% as the base case for Q4, the model assumes the 2020-end unemployment rate to be 10.3% and 15.0% as its low and high case assumptions, respectively. Projections are not reported on a monthly basis, so the ΔU will be calculated by subtracting scenario assumptions for the Q4 2020 unemployment rate by 3.5% (unemployment rate in December 2019).

	Low	**Base**	**High**
Est. Q4 Unemployment (a)	10.3 %	11.7 %	15.0 %
'19 Dec. Unemployment (b)	3.5 %	3.5 %	3.5 %
ΔU = (a) – (b)	6.8 %	8.2 %	11.5 %
Implied % Surprise*	102.8 %	122.28 %	168.15 %
Expected # of Data Breaches**	1704	1704	1704
Implied # of Data Breaches	3456	3788	4569

** Using the linear relationship derived in Section III, Figure 8*
*** Using the exponential relationship derived in Section II, Figure 3*

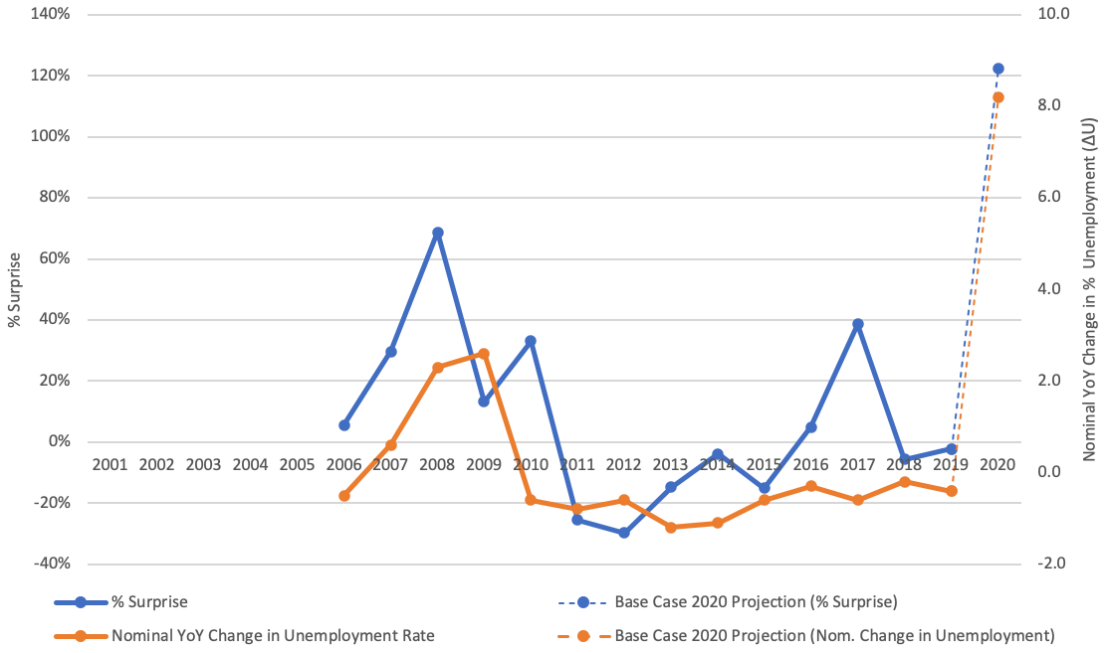
Figure 10: % Surprise as a function of YoY nominal change in unemployment rate (low, base, high case projections)



Source(s): Identity Theft Resource Center, St. Louis Fed

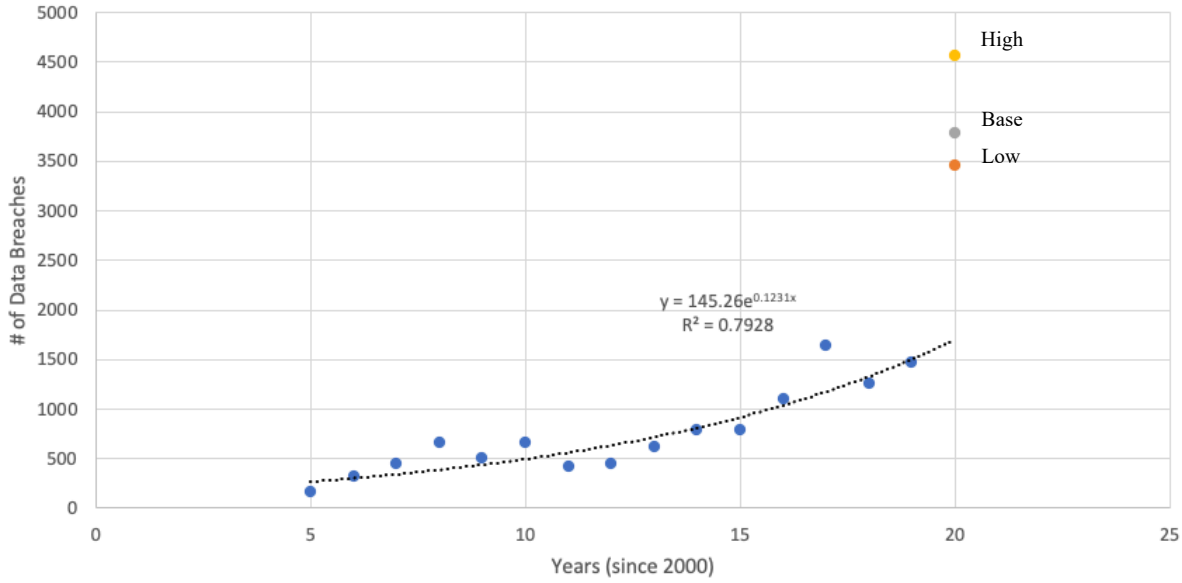
Cheaheon "Ian" Lim
ilim@college.harvard.edu

Figure 11: % Surprise versus YoY nominal change in unemployment rate with base case projections for 2020



Source(s): Identity Theft Resource Center, St. Louis Fed

Figure 12: Number of data breaches in U.S. over time (2000-2019) and estimates for 2020 based on scenario assumptions



Source(s): Identity Theft Resource Center

V. Concluding Remarks

If model fundamentals do end up holding for 2020, the GVC's potential implications on cybersecurity are daunting. As represented in Section IV (Figures 11 and 12), the nation may see a dramatic uptick in the number of data breaches and cyberattacks in the present year. Factors specific to the GVC, such as cybersecurity issues that arise from transitioning to remote work, may also create additional vulnerabilities that could translate to a higher number of data breaches.

However, this report is by no means comprehensive—it merely represents a preliminary data-based survey that I have conducted on the historical performance of the cybersecurity industry, which I then use to provide a rough estimate for what we may see in 2020 as a result of the GVC. As mentioned under Section III, the present situation is unprecedented in terms of the projections this report attempts to model for, meaning that model fundamentals may not end up holding for 2020. There are also a number of extraneous factors, such as a potential second wave of cases in the Fall, that may affect scenario assumptions.

Appendix

i. Constituents of Cybersecurity Basket (partially sourced from CIBR ETF)

Akamai Technologies (NASDAQ: AKAM)
A10 Networks (NYSE: ATEN)
Broadcom Inc (NASDAQ: AVGO)
Booz Allen Hamilton (NYSE: BAH)
Check Point Software Technologies (NASDAQ: CHKP)
CrowdStrike (NASDAQ: CRWD)
Cisco Systems (NASDAQ: CSCO)
CyberArk (NASDAQ: CYBR)
FireEye (NASDAQ: FEYE)
F5 Networks (NASDAQ: FFIV)
Fortinet (NASDAQ: FTNT)
Itron (NASDAQ: ITRI)
Juniper Networks (NYSE: JNPR)
Leidos (NYSE: LDOS)
ManTech International (NASDAQ: MANT)
Mimecast (NASDAQ: MIME)
MobileIron (NASDAQ: MOBL)
Cloudflare (NYSE: NET)
Okta (NASDAQ: OKTA)
OneSpan (NASDAQ: OSPN)
Palo Alto Networks (NYSE: PANW)
Proofpoint Inc (NASDAQ: PFPT)
Qualys (NASDAQ: QLYS)
Ribbon Communications (NASDAQ: RBBN)
Radware (NASDAQ: RDWR)
Rapid7 (NASDAQ: RPD)
Science Applications International Corporation (NYSE: SAIC)
SailPoint Technologies (NYSE: SAIL)
Splunk Technology (NASDAQ: SPLK)
Tenable Holdings (NASDAQ: TENB)
Tufin (NYSE: TUFN)
VMware (NYSE: VMW)
Varonis Systems (NASDAQ: VRNS)
Verisign (NASDAQ: VRSN)
Zix Corp (NASDAQ: ZIXI)
Zscaler (NASDAQ: ZS)