Characteristics of Bitcoin Users: An Analysis of Google Search Data

Matthew Wilson
Aaron Yelowitz*

Abstract: The anonymity of Bitcoin prevents analysis of its users. We collect Google Trends data to examine determinants of interest in Bitcoin. Based on anecdotal evidence regarding Bitcoin users, we construct proxies for four possible clientele: computer programming enthusiasts, speculative investors, Libertarians, and criminals. Computer programming and illegal activity search terms are positively correlated with Bitcoin interest, while Libertarian and investment terms are not.

JEL Codes: E42, F33, K42, K49
Keywords: Bitcoin, Digital currency, Google search data, Libertarians, Illegal Activity

* Yelowitz (corresponding author) Contact Information
E-mail – aaron@uky.edu
Telephone – 859-257-7634
Fax – 859-323-1920

Electronic copy available at: http://ssrn.com/abstract=2518603

Introduction

Bitcoin, a virtual global currency, has been the topic of much media, internet and policy discussion. Over 13.4 million bitcoins are in circulation and have a total market value of $4.6 billion.[1] Little is known about the characteristics of Bitcoin users, even though thousands of businesses accept bitcoins as payment. Transactions with Bitcoin are near anonymous due to the cost associated with identifying a user's electronic signature. Although some convenience sampling exists of Bitcoin enthusiasts, no systematic data collection has been done.

We use Google Trends (hereafter, "GT") data to study the clientele driving interest in Bitcoin, with the caveat that search query interest need not imply active participation. Based on anecdotal evidence about Bitcoin users, we construct proxies for four possible clientele: computer programming enthusiasts, speculative investors, Libertarians, and criminals. Illegal activity and computer programming are both positively associated with Bitcoin use, while no association exists for Libertarian ideology or investment motives in most specifications.

The Bitcoin Market

Bitcoin was created in 2009 as an unregulated, alternative method of exchange for online payments. Upon signing up for an account, an individual receives a electronic signature that secures transactions and disallows double spending (enforced by a diverse computer network). This process circumvents conventional methods that involve trust in and fees to a third-party. Conventional methods involve third-party fees, deterring small transactions (Nakamoto, 2008).[2] Anonymity is theoretically achieved due to Bitcoin's encryption, with the sole link being the electronic signature. Meiklejohn, et al. (2013)

[1] https://blockchain.info/charts/total-bitcoins
[2] https://bitcoin.org/bitcoin.pdf

find that anonymity is nearly impossible with large scale transactions, but there are high costs to identifying users.

Who Might Be Bitcoin Users?

Profit and politically-charged aspirations coincide with the basic design of the Bitcoin market. Prices for bitcoins have fluctuated enormously over time, which might prove tempting for a speculative investor. The unregulated set-up makes it appealing to Libertarians who philosophically oppose "inflationary central-bank meddling."[3] Other clientele appreciate Bitcoin's market structure for different reasons. For example, Bitcoin has appeal among computer programmers; miners (the term for those seeking to discover new bitcoins) can earn the currency in exchange for utilizing special software to authenticate real-time Bitcoin transactions.[4] The anonymity of Bitcoin is attractive for criminal activity. The October 2, 2013 FBI takedown of the Silk Road website – an online marketplace "for everything from heroin to forged passports" where transactions took place in bitcoins – highlighted the importance of Bitcoin's perceived anonymity and led to a 22% reduction in Bitcoin's price.[5]

In order to understand the underlying rationale for Bitcoin use, Lui (2013) surveyed 1,133 members of the Bitcoin community (by posting links on Bitcoin websites).[6] The survey identified three key motives: curiosity, profit, and political. Respondents (which included both owners and non-owners of Bitcoin) are likely unrepresentative of the larger community; for example, those using Bitcoin for illegal activity are unlikely to participate.

[3] http://www.economist.com/news/finance-and-economics/21599053-chronic-deflation-may-keep-bitcoin-displacing-its-fiat-rivals-money
[4] http://www.bitcoinmining.com/
[5] http://online.wsj.com/articles/SB10001424052702303722604579115692946177328 and https://www.tradingview.com/v/4xVX2cFq/
[6] http://simulacrum.cc/2013/04/13/overview-of-bitcoin-community-survey-feb-mar-2013/

GT Data

We collected GT search query data from January 2011 to July 2013 for all US states and Washington DC.[7] We looked for terms related to Bitcoin and its possible clientele.[8] Some of these correlations are inherently difficult to measure, due to the sensitivity of the activity; Stephens-Davidowitz (2013, 2014) argues, however, that Google data are unlikely to suffer from major social censoring, and uses GT to explore child abuse and racial animus.[9] Although it is conceivable that higher Bitcoin search volume need not translate into increased market participation, Kristoufek (2013) demonstrates a strong positive correlation between Bitcoin searches and exchange prices.

GT can be used to extract data for precise search terms and more general topics (see Figure 1). Search terms will return data for the exact query while topics count related searches too.[10] For instance, the topic "Bitcoin (Currency)" includes the terms "Bitcoin", "Bitcoins", "Bitcoin Mining", "Bit Coin", "Bitcoin exchange", "Bitcoin price" and "Bitcoin value". We use search topics for Bitcoin (under "Currency") and Computer Science (under "Discipline"). For other clienteles – Illegal Activity, Libertarians and Speculative Investors – we use the search terms "Silk Road", "Free Market", and "Make Money" respectively.[11]

GT does not report raw search counts for a topic; such counts would be misleading because Google's popularity (and search queries) grow over time.[12] Instead GT computes the number of topic searches relative to all searches, normalizes the series so the highest value is 100, and scales all other values

[7] We start in January 2011 because GT better measures state-level search activity from that point. We end in July 2013 because the "Silk Road" website – unknown to most of the public – was shut down soon after and made front-page headlines in national publications.
[8] GT data has been predictive of behavior in diverse economic markets including entertainment, labor, and housing (Hand and Judge, 2012; Askitas and Zimmerman, 2009; Varian and Choi, 2009; Wu and Brynjolfsson, 2013). It has also been used for detecting health patterns, including influenza outbreaks and Lyme disease cycles (Ginsberg, et al., 2009; Seifter et al., 2010; Carneiro and Mylonakis, 2009).
[9] He shows that cross-sectional state variation in GT is highly correlated with other data sources; for example the search rate for the word "God" explains 65% of the variation in the percent of a state's residents believing in God.
[10] https://support.google.com/trends/answer/4355000?hl=en
[11] We attempted to use alternative terms for these concepts (such as "Libertarian" or "Ron Paul" for Libertarianism), but search interest was either too sparse or had a strong political cycle.
[12] https://support.google.com/trends/answer/4365533?hl=en

relative to the highest. Figure 2 illustrates the Bitcoin time series in California, where popularity peaked in April 2013. For each state, we initially compute a 31-month time series for the relative popularity of Bitcoin and each clientele grouping.[13] We then use GT to measure relative state-level popularity of each search term for the full period and scale each state-series relative to the most popular state. During the observed timeframe, the states with the highest interest in Bitcoin were Utah, Oregon, California, Washington, Nevada, New Hampshire and Vermont (see Figure 3). We then rescale each state-specific time series by its geographic popularity. Thus, using California's value of 94 from the geographic Bitcoin comparison, the entire California time series would be rescaled to 0.94 of its original value.

Our outlined methodology presents us with two limitations. First, GT samples its database and computes the index based on that sample.[14] We observed slightly different values for the index by refreshing the webpage, even with the same restrictions. Although the overall conclusions are unlikely to change from sampling, this prohibits exact replication. Second, GT gives a value of zero if it cannot gather enough data.[15] We exclude state-month observations with missing values. While every index has missing values for particular months, some states returned a missing value in the cross-sectional analysis, which prevents rescaling of the state-specific time series. Delaware, North Dakota, and Wyoming were excluded as they had missing values for "Free Market" and/or "Silk Road." Out of 1,488 (48 states x 31 months) potential observations, our analysis uses 794 with non-missing values on Bitcoin, Computer Science, Free Market, Silk Road, and Make Money. The most populous states tend to have the fewest missing state-month observations.

[13] Some states and search terms had weekly activity (such as California's Bitcoin activity in Figure 2). In such cases, we computed monthly averages for all non-missing values, and then rescaled the series with a maximum value of 100.

[14] https://support.google.com/trends/answer/4355213?hl=en&ref_topic=4365599

[15] https://support.google.com/trends/answer/4355164?hl=en&ref_topic=4365531

Empirical Results

Follow Stephens-Davidowitz (2014), we normalize each search rate to its z-score and estimate the following specification:

(1) $$BITCOIN_{jt} = \beta_0 + \beta_1 X_{jt} + \delta_j + \delta_t + \varepsilon_{jt}$$

where $BITCOIN_{jt}$ is Bitcoin interest in state j in month t, X_{jt} are clientele interest, and δ_j and δ_t are state and time fixed-effects. Each state-month is weighted by state population in July 2011 and standard errors are corrected for non-nested two-way clustering at the state and time levels (Cameron, Gelbach and Miller, 2011). By including fixed effects in our fully-saturated specification, the impact of clientele association on Bitcoin is measured through differential within-state changes over time (Yelowitz, 1995).

Results for a variety of specifications are presented in Table 1. Columns (1)-(3) progressively include additional controls for state and time. The inclusion of both state and time fixed effects identifies interest in Bitcoin by exploiting within-state changes over time. In this specification, interest in computer science and Silk Road are both positively associated with interest in Bitcoin and are statistically significant at the 10% level. The interpretation of the specification in column (3) is the following: a one-standard deviation increase in computer science interest leads to a 0.13 standard deviation increase in Bitcoin interest, while a one-standard deviation increase in Silk Road interest leads to a 0.09 standard deviation increase in Bitcoin interest. Column (4) adds a “placebo clientele” – searches for the singer Miley Cyrus. Reassuringly, inclusion of this placebo variable neither changes any of the inferences on the other clientele, nor is the variable itself significant.

Columns (5)-(6) interact each clientele search term with average monthly Bitcoin prices. Profit motivated clientele – such as speculative investors – may find Bitcoin more intriguing when prices are high. However, we again observe a positive association between Bitcoin interest and our two clientele groups of computer programming enthusiasts and those possibly engaged in illegal activity (in the interaction term, not the main effect). The other clientele groups remain insignificant.

Columns (7)-(9) include the state-level monthly unemployment rate. Columns (7)-(8) show that the inferences on computer science and illegal activity are unchanged, but there is some evidence that Libertarian activity also drives interest in Bitcoin (although the specification including interactions with Bitcoin prices is insignificant). Higher unemployment rates are negatively associated with Bitcoin interest. Columns (10)-(11) estimate the model from 2012 onward (when Bitcoin was more popular), while column (12) estimates it for the 24 states with at least 20 monthly observations. In all cases, fluctuations in computer science and illegal activity continue to drive Bitcoin interest, as well as the business cycle.

Discussion

Although many commentators have speculated about motives for using Bitcoin, our study is the first to systematically analyze Bitcoin interest, including the interest of hard-to-observe clientele. We find robust evidence that computer programming enthusiasts and illegal activity drive interest in Bitcoin, and find limited or no support for political and investment motives.

References

Askitas, N., Zimmerman, K.F. 2009. "Google Econometrics and Unemployment Forecasting." *Applied Economics Quarterly,* 55(2), 107-120.

Cameron, A.C., Gelbach, J.B., Miller, D.L. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics,* 29(2), 238-249.

Carneiro, H.A., Mylonakis, E. 2009. "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks." *Clinical Infectious Diseases*, 49(10), 1557-1564.

Ginsberg, J., Mohebbi,M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.. 2009. "Detecting Influenza Epidemics using Search Engine Query Data." *Nature,* 457, 1012-1014.

Google Trends. www.google.com/trends.

Hand, C., Judge, G. 2012. "Searching for the Picture: Forecasting UK Cinema Admissions using Google Trends Data." *Applied Economics Letters,* 19(11), 1051-1055

Kristoufek, L. 2013. "Bitcoin meets Google Trends and Wikipedia: Quantifying the Relationship Between Phenomena of the Internet era." *Scientific Reports,* 3(3415), 1-7.

Lui. 2013. "The Demographics of Bitcoin (Part 1 Updated)." http://simulacrum.cc/2013/03/04/the-demographics-of-bitcoin-part-1-updated/

Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G.M., Savage, S. 2013. "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names." In *Proceedings of the 2013 conference on Internet measurement conference,* IMC '13, 127-140.

Nakamoto, S. 2008. "Bitcoin: A Peer-to-Peer Electronic Cash System." https://bitcoin.org/bitcoin.pdf

Seifter, A., Schwarzwalder, A., Geis, K., Aucott, J. 2010. "The Utility of 'Google Trends' for Epidemiological Research: Lyme Disease as an Example." *Geospatial Health,* 4(2), 135-137.

Stephens-Davidowitz, S. 2013. "Unreported Victims of an Economic Downturn," Working Paper. http://static.squarespace.com/static/51d894bee4b01caf88ccb4f3/t/51e22f38e4b0502fe211fab7/1373777720363/childabusepaper13.pdf

Stephens-Davidowitz, S. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data." *Journal of Public Economics,* 118, 26-40.

Varian, H.R., Choi, H. 2009. "Predicting the Present with Google Trends." Google Research Blog http://googleresearch.blogspot.com/2009/04/predicting-present-with-google-trends.html . Available at SSRN: http://ssrn.com/abstract=1659302

Wu, L., Brynjolfsson, E. 2013. "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales." Available at SSRN: http://ssrn.com/abstract=2022293 .

Yelowitz, A.1995. “The Medicaid Notch, Labor Supply and Welfare Participation: Evidence From Eligibility Expansions.” *Quarterly Journal of Economics*, 110(4): 909-939.

Figure 1
Google "Search term" versus "Topic (Currency)"
Source: Google Trends (www.google.com/trends).

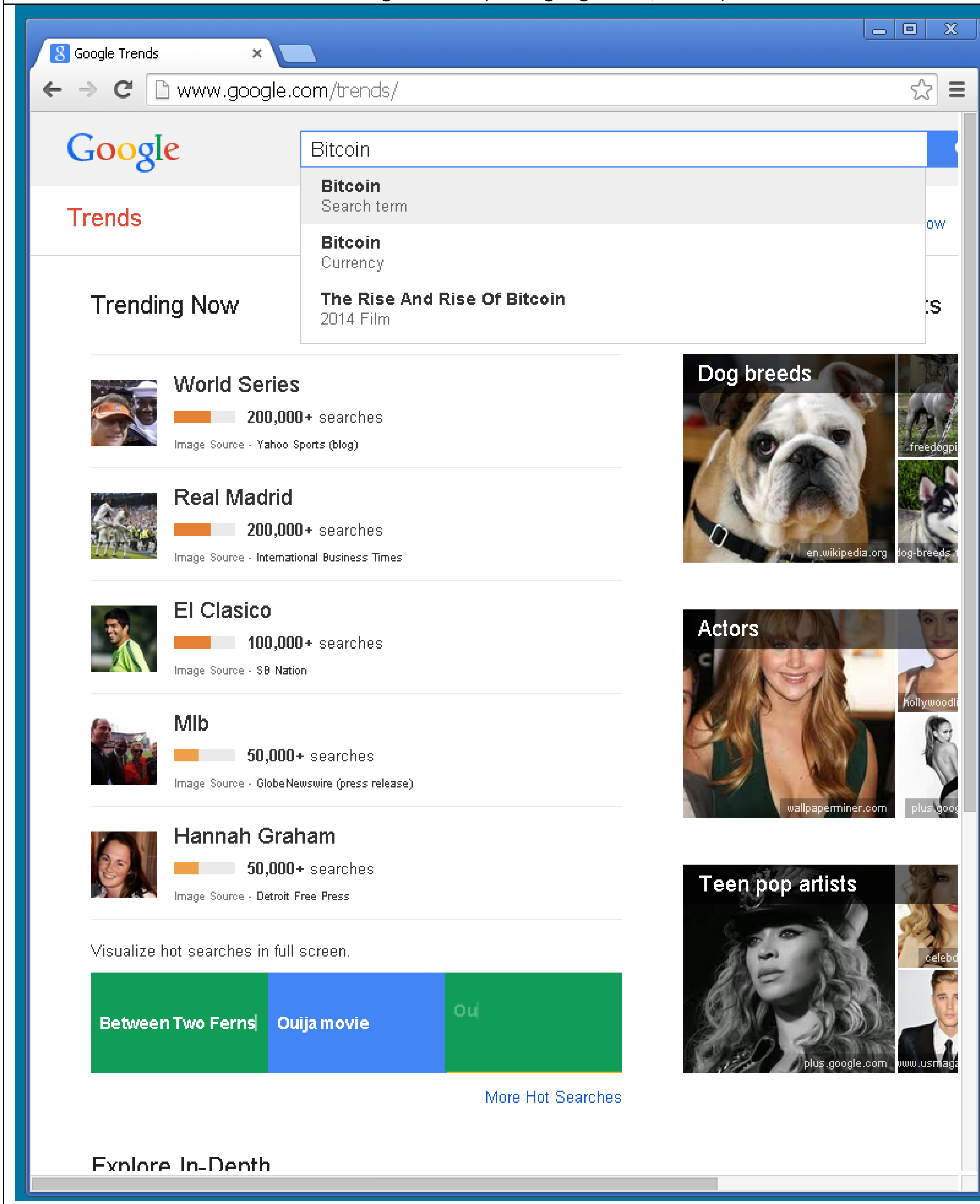

Figure 2
Index for Bitcoin Topic Search
California Time Series, January 2011-July 2013

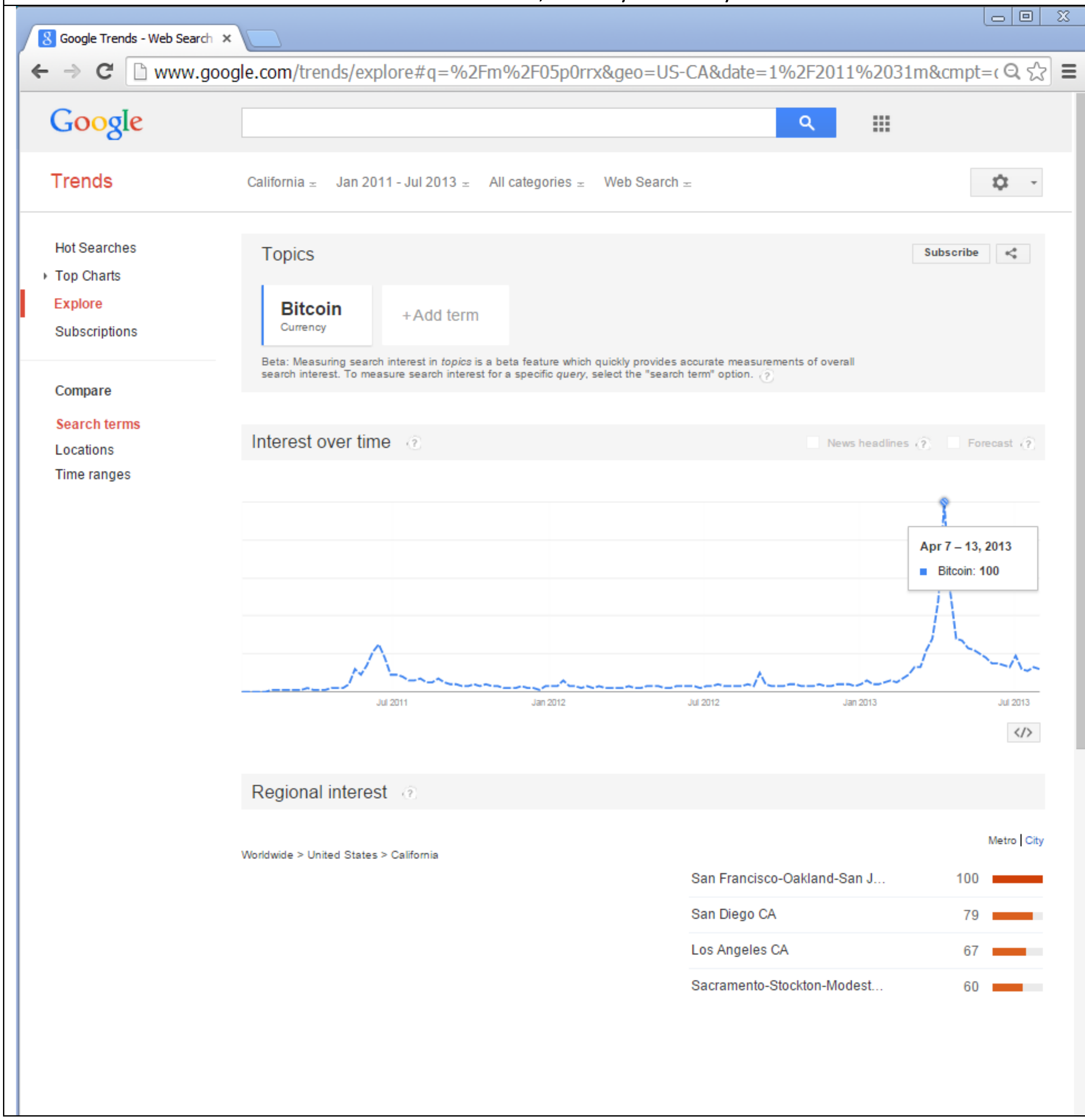

Figure 3
Index for Bitcoin Topic Search
Cross Sectional Popularity, January 2011-July 2013

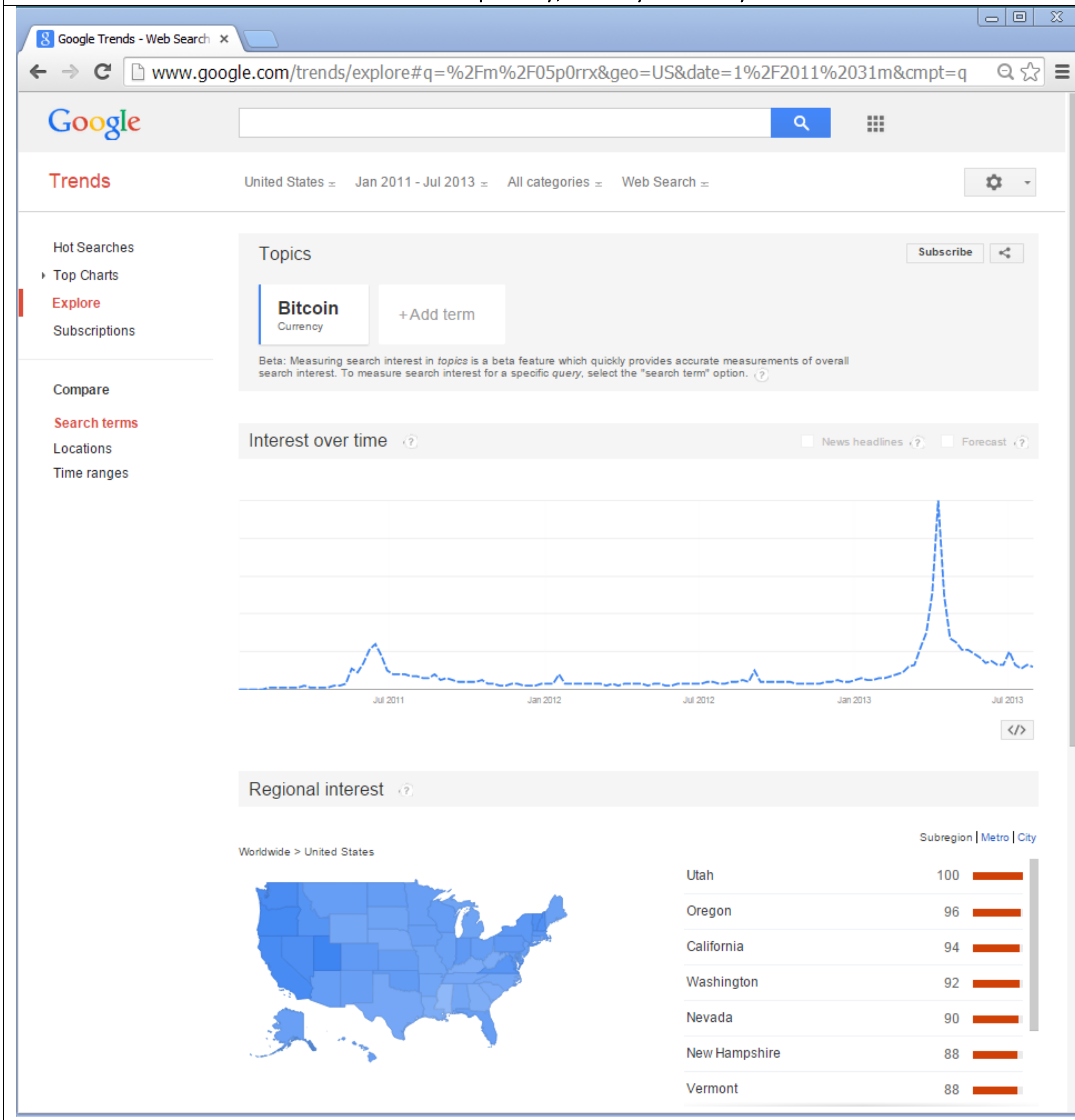

Table 1
Determinants of Bitcoin Search Interest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Computer Science	0.083	0.143	0.125	0.124	0.009	0.008	0.121	0.121	0.011	0.131	0.014	0.125
	(0.066)	(0.173)	(0.073)	(0.073)	(0.029)	(0.028)	(0.059)	(0.059)	(0.027)	(0.064)	(0.030)	(0.065)
Computer Science X PRICE/100					0.208	0.209			0.205		0.202	
					(0.068)	(0.068)			(0.064)		(0.062)	
Silk Road	0.948	1.080	0.093	0.093	-0.007	-0.007	0.076	0.076	-0.012	0.105	0.010	0.088
	(0.374)	(0.408)	(0.051)	(0.052)	(0.040)	(0.040)	(0.039)	(0.040)	(0.036)	(0.066)	(0.038)	(0.044)
Silk Road X PRICE/100					0.193	0.192			0.185		0.141	
					(0.101)	(0.100)			(0.097)		(0.082)	
Free Market	0.211	-0.172	0.023	0.023	-0.006	-0.005	0.031	0.031	0.003	0.036	0.004	0.021
	(0.076)	(0.058)	(0.022)	(0.022)	(0.022)	(0.021)	(0.019)	(0.019)	(0.019)	(0.025)	(0.021)	(0.020)
Free Market X PRICE/100					0.043	0.047			0.030		-0.011	
					(0.068)	(0.080)			(0.077)		(0.073)	
Make Money	0.052	0.085	-0.004	-0.004	0.004	0.004	0.005	0.005	0.016	-0.041	0.003	0.006
	(0.089)	(0.121)	(0.026)	(0.026)	(0.026)	(0.025)	(0.030)	(0.029)	(0.026)	(0.047)	(0.029)	(0.030)
Make Money X PRICE/100					-0.039	-0.045			-0.069		-0.095	
					(0.070)	(0.075)			(0.076)		(0.075)	
Miley Cyrus				0.021		0.031		0.015	0.034			
				(0.040)		(0.080)		(0.040)	(0.075)			
Miley Cyrus X PRICE/100						0.010			0.007			
						(0.115)			(0.101)			
Unemp. Rate							-0.121	-0.121	-0.080	-0.281	-0.203	-0.105
							(0.064)	(0.064)	(0.051)	(0.097)	(0.072)	(0.063)

Notes: Sample size is 794 in columns (1)-(9), 591 in columns (10) and (11) (2012 onward), and 580 in column (12) (states with ≥20 observations). Standard errors corrected for non-nested, two-way clustering at the STATE and MONTH levels. Observations weighted by population. State and time fixed effects included in columns (3)-(12). State fixed effects and a time trend included in column (2).