The Role of APIs in the Economy

Seth G. Benzell, Guillermo Lagarda, Marshall Van Allstyne

June 2, 2016

Abstract

Using proprietary information from a large percentage of the API-tool provision and API-Management industry, we explore the impact and role of APIs in the economies of the US and Europe. In this preliminary paper using incomplete data, we document what factors drive API adoption, and the relationship between API usage and firm income. APIs are developed by medium and large firms mostly in the Retail and Management industries. The number of APIs per firm increases by a factor of six from 2013 through 2015, while the amount of data flow per firm is roughly constant. We find evidence that firms that develop APIs have higher incomes in subsequent years than firms that do not. When the complete data set becomes available, we will further investigate the mechanisms for these relationships, as well as estimate the contribution of APIs to US GDP and skill biased technological change.

1 Motivation

The goal of this research is to understand the economic value of APIs and their role in changing the economy. APIs (application programming interfaces) are systems designed to ease the development of programs which interact with a firm's databases. API boosters claim many benefits from API adoption. These include: Increased code re-usability, lower barriers to entry for developers new to a firm's data, and lower information technology maintenance costs. These benefits should complement firm strategies that emphasize internal agility and ecosystem growth. Believers have made large investments in these systems: According to Programmable Web, the amount of Web APIs have increased from a few hundred in 2005 to over ten thousand today.

We investigate how investments in developing software and APIs makes firms valuable and productive. In particular, we want to answer four questions: What sort of firms adopt API technologies? What is the impact of API technology adoption on firm outcomes? Through what mechanisms are these effects realized? And finally, what are the general equilibrium and long-term consequences of API technology adoption?

According to the BLS, in 2014 there were 302,150 computer programmers in the US. Depending on how one classifies other occupations, there may be over three million workers engaged in writing code. However, the nature and magnitude of these economic contributions is poorly understood.

For example, microeconomic theory (Porter and Stern, 2000; Parker and Van Alstyne, 2016) argues for the importance of code spillovers. A recent study of SourceForge also offers evidence that spillovers exist (Eilhard and Ménière, 2009). However, many empirical questions about digital spillovers remain. For example, it remains an open question to what extent coding 'spillovers' are due to code accumulation versus programmers 'learning by doing'. As APIs are meant to complement and enhance the impact of normal programming, we hope that this research will also shed light on the broader question of code's impact in the economy.

2 Data Description

To conduct this analysis we will create a unique data-set matching API-tool providers' data on API production and use with firm level information on employment and outcomes.

For data on firm inputs and outputs, our primary sources are Compustat for North American publicly traded firms, and OSIRIS for European firms. These contain a wide range of firm characteristics, logged at the quarterly and yearly level. We are looking into supplementing this with publicly available individual level employment information.

For data on API development and usage, we are collecting information from API management firms. For all companies contracting with the API management firm, we are collecting the following at the monthly level:

- The date the company first signed a contract with the API management firm
- The list of APIs in use
- The name of the API and any qualitative information available about its role (e.g. HR management, inventory management, etc.)
- Whether the API is open to the public or closed and used primarily inside the firm
- The number of developers with permission to develop for the API or programs with permission to execute calls on the API
- The number of calls handled by the API
- Any information available on the complexity of the API or on the amount of time and resources invested in developing the API
- Characteristics of information requested in API calls (e.g. average file size, distribution of file types)
- General characteristics of the contract signed by the firm (e.g. do they pay by the call, or by the API? Have there been important changes in how contracts were written?)

In our preliminary data, we have the following panel variables available at the quarterly level

- Number of APIs
- Number of calls per API
- Volume of data retrieved per API

3 Characteristics of Firms Adopting APIs

Table 1 gives information on the types of firms in our preliminary data which have developed APIs. Table 3 gives this information for all Compustat firms.

The APIs in our data are widely dispersed in size and leverage. Figure 1 gives additional information on the characteristics of firms adopting APIs. Table 2 shows variance in API usage amongst firms that have at least one API in a year. Note that the first year of API observations is 2013.

Firms in our sample had an average of 5 APIs in 2013 with about 2.9 million total calls, while in 2015 the average firm had 32 APIs with about 3.2 million calls. Meanwhile the average amount of data per firm decreased slightly, meaning that APIs got smaller over time. This suggests that API development leads to greater API specialization rather than creation of one overall API for the entire firm.

4 Impact of API Adoption

We seek to understand how investments in software production, (e.g. the employment of heterogeneous programmers) lead to the evolution of assets (a firm-specific code package and open or closed APIs) which are in turn used as a factor of production?

We use three techniques to estimate $Y_i^T - Y_i^N = \gamma$, where Y income for i firm when T-treated (the firm has at least one API) or N-not treated. The results are given in table 2.

In the first and second specifications, we fit a linear regression with the listed controls. The second specification differs from the first in that the leverage of a firm is included as an additional covariate. These

Financial Summary Statistics, Firms Adopting APIs

Year	Variable	Firms	Mean	Std. Dev.	Min	Max
2010	Net Income	38	1434.68	3542.75	-224.16	19085.00
	Capital Investment	37	1297.06	4193.54	2.24	20302.00
	Leverage	37	1.47	6.37	-15.19	34.56
	Market Value	32	24118.85	41372.40	306.15	173667.73
	R&D % Income	23	5.69	8.41	0.00	26.70
	Operating Profits	37	3525.54	8430.69	13.75	38952.00
2011	Net Income	40	1050.25	1617.17	-83.02	8572.00
	Capital Investment	40	1269.24	4018.84	2.18	20272.00
	Leverage	38	0.84	4.72	-20.71	13.71
	Market Value	32	26364.90	43786.66	244.47	179217.72
	R&D % Income	25	5.57	7.97	0.00	25.42
	Operating Profits	40	3355.12	7166.68	-1.15	34686.00
2012	Net Income	41	1161.66	1906.26	-195.87	9019.00
	Capital Investment	40	1349.40	3968.17	2.50	19728.00
	Leverage	40	0.28	5.46	-30.52	10.83
	Market Value	33	27901.75	45647.53	297.86	188148.83
	R&D % Income	24	7.03	9.43	0.00	31.17
	Operating Profits	40	3289.89	6937.90	-10.76	31140.00
2013	Net Income	22	1919.89	3766.79	-536.87	18249.00
	Capital Investment	22	1905.17	4774.90	5.08	21228.00
	Leverage	22	2.20	3.05	0.12	16.97
	Market Value	22	31331.18	45231.01	516.41	183757.27
	R&D % Income	17	5.10	7.52	0.00	25.95
	Operating Profits	22	5478.64	12113.35	8.41	49374.00
2014	Net Income	33	1464.19	1754.84	-72.37	6224.00
	Capital Investment	33	2255.65	4975.84	17.21	21433.00
	Leverage	33	2.65	4.32	0.16	17.57
	Market Value	33	36009.49	46906.25	449.08	174228.41
	R&D % Income	33	5.14	9.03	0.00	26.87
	Operating Profits	33	4800.88	7379.46	31.60	31689.00
2015	Net Income	40	2297.67	3840.38	-43.21	13345.00
	Capital Investment	40	2210.31	5639.84	21.96	20015.00
	Leverage	40	3.73	8.28	0.20	31.16
	Market Value	40	42373.17	65740.92	771.83	211447.39
	R&D % Income	40	6.50	11.15	0.00	28.87
	Operating Profits	40	6487.03	12952.94	23.06	47845.00

Table 1: Financial Summary Statistics for firms that adopt APIs during or after 2013. Variables are in millions, except for leverage ratio and sales. Sales are quarterly sale quantities indexed by earliest period on record (Observation to be replaced with firm). Net income is operating and non-operating income minus non-extraordinary expenses. Operating profit is operating income minus operating expenses. Observations are quarterly, number of firms for a variable are those with at least one quarter of data in a year.

Main Characteristics of Firms that Adopted APIs			
	2013	2014	2015
Main Sectors			
Retail (NAICS 44-45)	67.6%	72.10%	70.20%
Management (NAICS 55-56)	17.4%	20.20%	19.10%
Information & Financial (NAICS 51-52)	7.0%	5.30%	4.60%
Average Others	0.4%	0.70%	0.63%
Region (Publicly Traded in USA only)			
USA	83.4%	80.0%	79.3%
EUROPE	10.1%	9.2%	10.2%
AMERICAS	4.4%	8.1%	7.7%
ASIA	2.0%	2.3%	2.0%
Proximity to Provider (Euclidean L2)	31.1	30.98	27.3
USA (firms based in USA)	31.1	30.98	27.3
EUROPE (firms based in EU)	22.7	21	20.4
Size			
Employees	674.22	1180.8	943.2
Margin after Operational Expenditures	11.80%	14.80%	18.60%

Figure 1: Characteristics of firms that adopt APIs during or after 2013, by year.

specifications yield the result that firms that adopt APIs have significantly higher incomes in subsequent years.

Of course, future income is unlikely to depend on sales, capital expenditure, and leverage in a linear way. Therefore, our second approach is to generate matching estimators. These are based on imputing counterfactual outcome value for each firm. The matching estimator takes the form:

$$\tau = N^{-1} \sum [\hat{Y_i^T} - \hat{Y_i^N}]. \tag{1}$$

Specification 3 uses propensity scoring to estimate the effect of API adoption. Our current preferred specification (the fourth), nearest neighbor matching, uses the firms that are most similar ex-ante to the the API using firms to estimate the effect of API use. Once APIs are adopted by a firm they have significantly more income than firms that are ex-ante similar. Our point estimate is that their income is over a third of a million dollars higher every year.

Tables 4, 5, 7, 6, 8 , and 9 linearly estimate the effect of different measures of API use on income. The results again are optimistic: Volume and Traffic seem to have a positive effect on revenues and operating profits. An interesting finding is that after adoption of API, using more programmers is associated with lower revenues. While the analysis still a work in progress, this might offer a hint of interesting questions regarding the effect (efficiency) that an API might have on a single programmer.

The estimates of **??** and 2 are likely to be biased. API usage is probably caused by unobservable firm characteristics, like anticipated demand growth or general firm tech-savvyness, that might be correlated with future income. In future work we will use IV and other approaches to eliminate this bias.

As more information becomes available, we would also like to understand how APIs make their impact. For example, to what extent do code and APIs complement or substitute for other factors of production? How do they differ? To what extent does the openness of an API matter to its function in a firm? Does the

API Usage Statistics

Year	Variable	Firms	Mean	Std. Dev.	Min	Max
2013	Number of APIs	22	5.00	5.67	1	25
	Developers	22	80.26	148.06	1	492
	Calls	22	2.87	12.20	2	57.40
	Data	22	126000.00	522000.00	0	2340000.00
	Data per Call	22	43971.37	42786.89	2	40766.55
2014	Number of APIs	33	15.52	18.86	1	70
	Developers	33	329.27	888.33		4777
	Calls	33	1.20	2.55	2	10.80
	Data	33	26700.00	104000.00	0	571000.00
	Data per Call	33	22304.14	40776.61	2	52870.37
2015	Number of APIs	40	32.05	59.37	1	331
	Developers	40	365.05	598.69	1	3171
	Calls	40	3.23	11.10	1	66.60
	Data	40	35700.00	111000.00	0	546000.00
	Data per Call	40	11043.89	10000.00	2	8198.20

Table 2: API Usage Summary Statistics for firms that adopt APIs during or after 2013. Data is volume flow in million bytes, calls are in millions. Developers is number of developers authorized to write programs for a firm's APIs. Observations are quarterly, and number of firms is the number of firms with at least one observation in a year.

function of APIs as a factor of production vary by industry? Firm strategy? What kind of investments have the highest return? Investments in open APIs, closed APIs, or investments in bespoke programs? Is there an inter-temporal tradeoff?

As another method of understanding the value and mechanisms by which APIs and related investments impact firms, when our final data set becomes available we will estimate the following production functions:

$$Y_{i,t} = f_{Y,i}(\overrightarrow{A_{m,t}}, P_t, X_{Y,t})$$

$$P_t = f_{i,P}(P_{t-1}, \overrightarrow{A_{m,t}}, X_{P,t})$$

$$A_{m,t} = f_{i,A}(A_{m,t-1}, X_{A,m,t})$$

Where Y_i is some characteristic of a firm in industry i (such as revenue), A_m is the (unobservable) quality in efficiency units of every API m, and P is the quality of the firm's private program stock. X are any other inputs used in the production of these. These flexible functions allow for many types of spillovers, including learning by doing, code-reuse (and depreciation), and increasing returns to scale.

We will estimate these equations using the following approaches. First, we will try to discover a proxy (which may be a compound of such factors as API calls by type) to use in the place of unobservable A_m and P. We hypothesize that APIs with different functions (e.g. sales, login, inventory) and orientations (e.g. B2B, B2C, internal) will differentially impact a firm's inputs and outputs. A second approach to estimating $\hat{A_m}$ and P_t is to perform a two-step procedure to generate an in sample forecast series.

For each of these approaches we will then use the Generalized Method of Moments estimator of Arellano and Bond (1991), also known as difference GMM, and an augmented version (system GMM) developed by Arellano and Bover (1995).

These estimators are designed for dynamic panel models with: a shorter time dimension, and large number of observations (in our case firms or code modules); a linearized functional relationship; a left-hand-side variable that is dynamic, depending on its own past realizations; independent variables that are not strictly exogenous, meaning that they are correlated with past and possibly current realizations of the error;

Treatment Effect Estimation
Outcome Variable: Income (Revenues before Extraodinary Events)
Treatment Variable: Adoption of API (Binary, adjusted in time)

	Regression (IPW)		Propensity	Nearest
	Baseline	Leverage	Match	Neighbor
ATE	162.4634**	187.8349*	492.1583**	364.7665**
	(13.34328)	*(14.76836)*	*(125.6853)*	*(69.51876)*
Sales	0.05235+	0.052296*		
	6.20	*6.20*		
CAPX	-0.0564	-0.0563152+		
	(6.19)	*(6.37)*		
lev		-0.0461+		
		(7.745)		
Constant	27.56532**	31.28339**		
	(13.01)	*(17.01)*		

Std. Errors statistics in parentheses
+ p<0.10, * p<0.05, ** p<0.01

Figure 2: Estimation of the effect of API technology adoption under different specifications. Outcome is the net income variable described above, measured in millions. Sales is total firm sales in quantities indexed by sales in the firm's first year of observation.

fixed individual effects; and heteroskedasticity and autocorrelation within individuals and possibly across them. Robustness to autocorrelation is especially important given that we believe A_m and P accumulate over time. These GMM estimators take the first difference of the estimated equation to eliminate the fixed effects term and then use the lagged value (or future value in the forward orthogonal case) of the right hand side variables as instruments to estimate the coefficients.

5 General Equilibrium Impact

A large body of macroeconomic theory (Jones, 2002, 2005; Romer, 1986; Furman et al., 2002; Benzell et al., 2016) argues for the ongoing importance of code and code spillovers. The skill biased technological change literate emphasizes the importance of these and related technical changes on wages.

In future work, we would also like to estimate the overall impact of APIs and related code investments on the economy. In particular we plan to make an estimate of APIs' overall economic contribution and average RTI. Finally, with a rich enough data set, we hope to explore the contribution of APIs to skill biased technological change.

References

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics 68*(1), 29–51.

Benzell, S., L. Kotlikoff, G. Lagarda, and J. Sachs (2016). Robots are us: Some economics of human automation. *NBER Working Paper*.

Eilhard, J. and Y. Ménière (2009). A look inside the forge: Developer productivity and spillovers in open source projects. *Available at SSRN 1316772*.

Furman, J. L., M. E. Porter, and S. Stern (2002). The determinants of national innovative capacity. *Research policy 31*(6), 899–933.

Jones, C. I. (2002). Sources of us economic growth in a world of ideas. *American Economic Review*, 220–239.

Jones, C. I. (2005). Growth and ideas. *Handbook of economic growth 1*, 1063–1111.

Parker, G. and M. W. Van Alstyne (2016). Innovation, openness, and platform control. *Management Science Forthcoming*.

Porter, M. E. and S. Stern (2000). Measuring the "ideas" production function: Evidence from international patent output. Technical Report Working Paper #7891, National Bureau of Economic Research.

Romer, P. (1986). Increasing returns and long-run growth. *The Journal of Political Economy 94*(5), 1002.

Financial Summary Statistics, All Compustat Firms

Year	Variable	Firms	Mean	Std. Dev.	p25	p50	p75
2010	Net Income	600	165.02	1124.24	-4.00	3.47	51.27
	Capital Investment	544	158.44	899.27	0.34	5.24	45.80
	Leverage	599	12.39	710.53	0.27	0.93	2.61
	Market Value	502	2717.58	13778.56	33.57	212.69	1139.54
	R&D % Income	259	270.46	3932.85	0.13	3.65	15.65
	Operating Profits	548	496.47	3191.34	-0.13	22.52	178.61
2011	Net Income	603	180.23	1291.11	-3.44	4.34	60.48
	Capital Investment	547	184.84	1043.71	0.33	6.13	56.21
	Leverage	602	12.29	567.50	0.26	0.95	2.68
	Market Value	500	2815.08	15005.92	32.54	191.44	1169.03
	R&D % Income	258	836.26	20578.36	0.11	3.73	16.39
	Operating Profits	551	531.78	3324.86	-0.20	24.67	198.13
2012	Net Income	623	177.59	1349.31	-4.34	3.88	57.80
	Capital Investment	566	200.67	1122.96	0.27	5.64	59.60
	Leverage	622	14.86	618.78	0.23	0.93	2.75
	Market Value	504	3073.15	17587.85	32.45	208.51	1274.91
	R&D % Income	266	775.72	17791.76	0.09	4.00	18.43
	Operating Profits	570	535.02	3457.00	-0.53	22.00	196.21
2013	Net Income	633	218.83	1865.12	-5.02	3.69	60.08
	Capital Investment	574	208.66	1203.91	0.19	5.20	60.12
	Leverage	633	4.82	303.45	0.20	0.89	2.56
	Market Value	516	3851.48	19087.22	40.90	288.37	1707.91
	R&D % Income	268	395.70	4378.62	0.16	4.10	18.92
	Operating Profits	579	571.46	3651.76	-0.91	19.55	197.32
2014	Net Income	597	199.04	1302.81	-5.48	5.12	74.80
	Capital Investment	544	232.65	1230.99	0.24	6.26	68.20
	Leverage	596	2.87	41.47	0.26	1.00	2.68
	Market Value	526	4099.40	20474.84	50.38	323.39	1839.94
	R&D % Income	256	551.07	7470.75	0.20	4.24	18.53
	Operating Profits	548	550.10	3023.48	-1.09	25.71	235.98
2015	Net Income	559	285.36	1856.41	-1.41	19.21	144.26
	Capital Investment	419	335.18	1437.76	1.78	24.78	129.69
	Leverage	556	2.85	13.32	0.45	1.20	3.11
	Market Value	534	6863.46	28674.46	139.96	917.40	3629.67
	R&D % Income	253	176.33	3152.15	0.00	4.20	15.97
	Operating Profits	512	725.87	3282.37	3.94	88.91	431.00

Table 3: Financial Summary Statistics for firms that adopt APIs during or after 2013. Variables are in millions, except for leverage ratio. Net income is operating and non-operating income minus non-extraordinary expenses. Operating profit is operating income minus operating expenses. Observations are quarterly, number of firms for a variable are those with at least one quarter of data in a year.

Table 4: Linear Regression of Log Net Income, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Data	6.669^{+}	7.455^{+}	8.102^{+}	6.646^{+}
	(1.98)	(1.93)	(2.43)	(2.27)
Capital Expenditures (log)		-0.146	-0.409	0.0215
		(-0.55)	(-1.48)	(0.06)
Employees			0.00866	-0.00582
			(1.67)	(-0.59)
Leverage Ratio				0.254
				(1.65)
Constant	6.421^{**}	7.310^{**}	7.454^{**}	7.251^{*}
	(186.82)	(4.51)	(5.35)	(5.76)
Observations	720	720	720	680
R^2	0.394	0.429	0.663	0.823

t statistics in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 5: Linear Regression of Log Net Income, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Calls	6.448^{+}	7.430^{+}	8.125^{+}	6.695^{+}
	(1.95)	(1.94)	(2.48)	(2.36)
Capital Expenditures (log)		-0.170	-0.439	-0.00482
		(-0.63)	(-1.58)	(-0.01)
Employees			0.00877	-0.00571
			(1.71)	(-0.60)
Leverage Ratio				0.254
				(1.69)
Constant	6.607^{**}	7.677^{**}	7.920^{**}	7.544^{*}
	(205.71)	(4.51)	(5.45)	(5.75)
Observations	800	800	800	760
R^2	0.387	0.432	0.672	0.832

t statistics in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 6: Linear Regression of Log Net Income, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Developers	-0.0385^{+}	-0.0448^{+}	-0.0563*	-0.0467^{+}
	(-1.91)	(-1.91)	(-3.39)	(-2.89)
Capital Expenditures (log)		-0.175	-0.547^{+}	-0.182
		(-0.64)	(-2.33)	(-0.55)
Employees			0.0111^{+}	-0.000597
			(2.59)	(-0.07)
Leverage Ratio				0.197
				(1.44)
Constant	6.797**	7.933**	8.489**	8.133**
	(76.46)	(4.47)	(6.91)	(6.60)
Observations	800	800	800	760
R^2	0.377	0.424	0.785	0.873

t statistics in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 7: Linear Regression of Log Operating Profits, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Data	0.237	0.272	0.319	0.236
	(1.15)	(1.13)	(1.80)	(1.64)
Capital Expenditures (log)		-0.00643	-0.0257	-0.00115
		(-0.39)	(-1.75)	(-0.07)
Employees			0.000633^{+}	-0.000191
			(2.30)	(-0.40)
Leverage Ratio				0.0145
				(1.91)
Constant	0.172**	0.207^{+}	0.222*	0.191^{+}
	(87.31)	(2.09)	(3.04)	(3.08)
Observations	800	760	760	720
R^2	0.181	0.205	0.658	0.845

t statistics in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 8: Linear Regression of Log Operating Profits, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Calls	0.229	0.272	0.322	0.241
	(1.14)	(1.14)	(1.85)	(1.73)
Capital Expenditures (log)		-0.00731	-0.0269	-0.00226
		(-0.44)	(-1.83)	(-0.13)
Employees			0.000638^{+}	-0.000184
			(2.35)	(-0.39)
Leverage Ratio				0.0144
				(1.95)
Constant	0.176**	0.219^{+}	0.240*	0.201^{+}
	(95.23)	(2.10)	(3.15)	(3.11)
Observations	880	840	840	800
R^2	0.177	0.207	0.666	0.853

t statistics in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 9: Linear Regression of Log Operating Profits, with Firm Fixed Effects

	(1)	(2)	(3)	(4)
Developers	-0.00153	-0.00183	-0.00261*	-0.00208*
	(-1.28)	(-1.30)	(-3.72)	(-4.31)
Capital Expenditures (log)		-0.00845	-0.0338*	-0.0135
		(-0.52)	(-3.39)	(-1.38)
Employees			0.000755*	0.000104
			(4.18)	(0.39)
Leverage Ratio				0.0110^{+}
				(2.69)
Constant	0.183**	0.234^{+}	0.276**	0.241**
	(37.22)	(2.24)	(5.35)	(6.52)
Observations	880	840	840	800
R^2	0.214	0.254	0.861	0.959

t statistics in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$