

Temporal Issues in Replication: The Stability of Centrality-Based Advantage

Yuan Shi,^a Olav Sorenson,^b David M. Waguespack^a

a) University of Maryland; b) Yale University

Abstract: The results of archival studies may depend on when researchers analyze data for at least two reasons: (1) databases change over time and (2) the sampling frame, in terms of the period covered, may reflect different environmental conditions. We examined these issues through the replication of Hochberg, Ljungqvist, and Lu's (2007) research on the centrality of venture capital firms and their performance. We demonstrate (1) that one can reproduce the results in the original article if one uses data downloaded at roughly the same time as the original researchers did, (2) that these results remain fairly robust to even a decade of database updating, but (3) that the results depend sensitively on the sampling frame. Centrality only has a positive relationship to fund performance during boom periods.

Keywords: centrality; replication; status; venture capital

Much of the recent attention to the need for replication has focused on confirming the results of experiments, particularly in medicine and in psychology (e.g., Ioannidis 2005; Simmons, Nelson, and Simonsohn 2011). Replication often matters in these settings because the high cost of doing certain sorts of research restricts these experiments to small samples, raising the potential for false positives both through random chance and through the intentional and unintentional decisions of the researchers (Simmons et al. 2011).

By contrast, replication in sociology and other nonexperimental social science research might seem less important. Research in these fields typically involves regressions run on large, archival datasets. These studies have plenty of statistical power, reducing both the importance of the choices made by researchers and the odds of misattributing noise to being a signal. The data also often appear relatively immutable. Replication, therefore, would seem to involve little more than understanding the sensitivity of the results to the construction of variables and to estimation choices (e.g., Breznau 2015; Hannon and Defina 2016).

But archival studies have their own sources of vulnerability. Databases, particularly those regularly updated by government agencies or commercial vendors, change as additional data become available and as errors get corrected. Sampling methods and questions may change subtly across years. Social processes and market forces also evolve and drift over time, meaning that results that hold in one longitudinal sample may not in another that ostensibly covers the same population over a slightly different time frame.

Our replication explores the importance of these temporal issues—when the researcher engaged in the research and the period that he or she studied—through the replication and extended analysis of one of the most highly cited articles relat-

Citation: Shi, Yuan, Olav Sorenson, and David M. Waguespack. 2017. "The Stability of Centrality-Based Competitive Advantage." *Sociological Science* 4: 107-122.


Received: October 17, 2016

Accepted: December 6, 2016

Published: January 30, 2017

Editor(s): Delia Baldassarri

DOI: 10.15195/v4.a5

Copyright: © 2017 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

ing firm performance to centrality in an interorganizational network: Hochberg, Ljungqvist, and Lu's 2007 article in the *Journal of Finance* (henceforth abbreviated as HLL2007 in the interest of efficient exposition).

Like hundreds of other studies (e.g., Podolny 2001; Sorenson and Stuart 2001; Trapido 2007), HLL2007 relied on data on venture capital (VC) investments drawn from the database originally developed by Venture Economics and subsequently acquired and incorporated into the ThomsonONE product. HLL2007 found that VC firms that occupied central positions in the coinvestment network—regardless of the specific centrality measure used—invested in firms that enjoyed much higher rates of “exit” events (an initial public stock offering or a sale to another firm), a close correlate of the returns that the VC firms could expect from their investments (Phalippou and Gottschalg 2009). They also found that the companies in which central firms invested survived longer. As one of the early studies connecting the centrality of firms to their financial performance, it has garnered a great deal of attention, with over 200 citations to date in Thomson's ISI database and more than 1,000 in Google Scholar.

To explore the stability of centrality-based advantage in VC, we began by attempting to replicate the results found in HLL2007. Using the same source data and following their procedures for sample selection and variable construction, we produced three alternative replication datasets. Those datasets vary in terms of (1) the date of downloading the source data, either September 2004 or October 2013, and (2) the study period, either 1980–1999, as in the original study, or 1980–2008, in an extended replication. We then analyzed those samples using the procedures outlined in HLL2007, focusing on VC fund performance and portfolio company survival to follow-on funding. Overall, we could generally reproduce the results reported in HLL2007, finding only minor differences between the original study and the three replication datasets.

We next examined the fidelity of the underlying data, documenting that, for the 1980–1999 period, substantial retrospective modifications had been made to the source data from the 2004 download to the one in 2013. These modifications, however, had little effect on the correlation between centrality and performance, perhaps because most of the modified cases appear to relate to more marginal VC firms and their portfolio companies. Still, two researchers ostensibly studying the same time period in the same data source could easily find themselves using datasets with as little as 60 percent overlap. Other relationships, therefore, might end up being more sensitive to these changes. These retrospective revisions may also have been systematic, such as in the case of securities analysts modifying the I/B/E/S data to appear better at predicting the future (Ljungqvist, Malloy, and Marston 2009).

We finally explored the sensitivity of the positive correlation between centrality and performance to the sampling frame, exploring how the results of HLL2007 might vary across different periods: 1980–2008 (the full range of data available today), 1980–1990 (the early growth of the industry), 1991–1994 (a shakeout period), 1995–2001 (a boom period), and 2002–2008 (the post-boom bust). The results depend sensitively on the sampling frame: the positive effects of network centrality appear to hold only during the boom periods associated with the 1980s and the late 1990s.

By contrast, during the bust periods, the results suggest that centrality might even become a liability.

Temporal Issues in Replication

Absent some sort of error, the reader would generally expect that another social scientist with access to the same database would find the same results. But this assumption depends crucially on the stability of the database over time. These databases may nevertheless change over time for at least three reasons. The first has to do with the fidelity of the data itself.

Commercial database providers, such as Thomson Reuters, frequently modify their records over time. In some instances, they add cases as their researchers uncover additional data. In other cases, they correct errors. In other instances, they may update the information, for example by changing the address of a company. In most cases, however, they only maintain a single record and therefore the researcher cannot determine whether the case has always been there and has always had the same values or whether it changed at some point along the way. These changes therefore pose a challenge to scholars working with versions of the database accessed at different points in time. Although they believe themselves to be studying the same sample, they may unwittingly analyze distinct populations.

If these changes arise from actors with interests in the information in the database, they may prove especially problematic. Ljungqvist et al. (2009), for example, discovered that entries in the I/B/E/S database, which tracks the predictions of stock analysts, had been retrospectively altered so that analysts would appear to have made fewer errors than they actually did in predicting the future price movements of stocks. The firms employing these analysts have an incentive to rewrite the past because investors use the database to select investment advisors and managers. Those changes obviously bias any studies attempting to use the database to study analyst accuracy, but they may more subtly affect other studies using the database as well.

The second concerns the data generating processes. Both commercial and noncommercial collectors of data sometimes modify their methods over time. They may, for example, change the wording of the questions used to elicit particular responses. They might alter the definition of the sample. Or, they may choose to collect the data in a somewhat different manner. The VC data, for example, had originally been collected through a fax-based survey but has more recently been conducted by email.

Even when these changes have been well documented and seem innocuous, they may have consequences for research. Lee and Bearman (2016), for example, demonstrated that the shift in interviewing for the General Social Survey in 2004, from the spring to the fall, produced as an artifact the apparent increase in social isolation identified by McPherson, Smith-Lovin, and Brashears (2006). In the fall of an election year, respondents more probably interpreted the “talking about important matters” question as being about politics rather than about other aspects of their lives.

The third has to do with the implicit temporal sampling frame. Researchers using longitudinal data frequently analyze the entire range of data for which they have consistent variables. Sometimes they might truncate the earliest dates in the interest of data quality or relevance, but rarely do analyses either explicitly examine the stability of a process over time or define the time period under study on the basis of meeting some set of theoretical scope conditions. Yet, social processes evolve and drift over time. Results that hold in one longitudinal sample may not hold in another that covers the same population over a somewhat different period.

We explored these issues in the context of Hochberg et al. (2007), taking advantage of the fact that we have been collecting snapshots of the database at different points in time. From both an empirical and a theoretical point of view, one would like to have a better sense of how sensitive these results may have been to the time at which the researchers engaged in their analysis.

Hochberg et al. (2007)

VC firms invest in young companies that have a high potential for growth. Although they represent but a small portion of the money invested in start-ups, the fact that they focus on high-growth companies means that they play a disproportionately large role in job creation and economic growth. Samila and Sorenson (2011), for example, found that the median VC investment in the United States led to the creation of more than 300 jobs in the region hosting the start-up.

One of the notable features of this “market” has been the high levels of uncertainty involved and the importance of private information. Start-ups fail at extremely high rates for a variety of reasons. Sometimes they cannot successfully develop a product or build the organization needed to deliver it. Sometimes they find that people do not want their product as much as they had anticipated or that customers will not pay as much as it costs them to produce it. Sometimes they encounter intense competition from incumbents or other start-ups. In trying to assess whether an entrepreneurial team has a good idea, whether it has the ability to see the idea to fruition, and whether it will operate effectively and reliably, venture capitalists rely heavily on their personal relationships for information and insight about the individuals involved.

Given the importance of these relationships to screening potential investments, Sorenson and Stuart (2001) argued that VC firms should benefit from being in central positions in the industry. These positions afford VC firms greater reach, allowing them to consider a wider range of investments. Consistent with this expectation, they found that central VC firms invested in companies located further from them geographically and in companies across a wider range of industries.

Another notable feature of VC has been that these investors play very active roles in the companies in which they invest. They frequently advise the founder on decisions related to strategy and managing the firm. They often help to recruit key employees to the company. They may even play a role in connecting the company to buyers and suppliers. Sørensen (2007) estimated that these post-investment activities might account for as much as two-thirds of the variation across VC firms in the performance of their investments.

Given this active role, one might expect that entrepreneurs would have a strong interest in being backed by the best VC firm possible. Indeed, received wisdom among entrepreneurs suggests that “It isn’t getting the money [that’s important], it’s who the money comes from” (Sorenson and Stuart 2001:1554). Perhaps not surprisingly then, Hsu (2004), in evaluating how entrepreneurs chose among competing investment offers, found that entrepreneurs would give up a larger share of ownership in their company in exchange for smaller cash infusions from high-status VC firms. In other words, entrepreneurs would willingly pay a price to associate with a higher-status investor.

Consistent with these mechanisms, Hochberg et al. (2007) found that VC firms in central positions in the coinvestment network (1) experienced successful exits from a larger proportion of their portfolio companies and (2) invested in companies that survived longer. That might reflect either their more extensive contacts, allowing them to select from a larger pool of potential investments, or their status, allowing them to get better deals in these transactions. In either case, these theoretical mechanisms would point to Bonacich (eigenvector) centrality as the most appropriate measure of centrality (Bonacich 1987; Podolny 1993; Sorenson and Stuart 2001). HLL2007 nevertheless examined this relationship using a broad array of centrality measures: degree (total unique direct ties for firm i), indegree (the number of firms who invite firm i to join a syndicate), outdegree (the number of firms invited to join syndicates by firm i), eigenvector centrality (firm i ’s degree weighted by the quality of each tie’s connections), and betweenness centrality (the extent to which firm i serves as a stepping point on the shortest path between two others). They found relatively similar results across all five measures.

Replication Methodology

To assess the issues of data source fidelity and effect stability over time, we constructed three alternative replication datasets. In all cases we used records from ThomsonONE (formerly, Venture Economics and then VentureXpert). For the first replication dataset, labeled REP1, we selected records for the 1980–1999 period, with the outcomes measured through November 2003, for data downloaded in September 2004. Given that Hochberg et al. (2007) calculated dependent variables through November 2003 and that the first version of their working paper appeared in December 2004, we assumed that they downloaded their data sometime between December 2003 and March 2004.

The second replication dataset, labeled REP2, again studies the 1980–1999 period but draws data from an October 2013 download. The third replication, labeled REP3, studies the 1980–2008 period, with the outcomes measured through November 2012, based on data downloaded in October 2013.

Hochberg and her collaborators essentially regressed the performance of portfolio companies on the centrality of the VC firms investing in them. They used five different measures of centrality, but because Bonacich centrality seems the theoretically most appropriate measure, our tables focus on this construct. Unless otherwise noted we found consistent results with the unreported alternative centrality measures.

HLL2007 computed centrality based on annual adjacency matrices using a five-year moving window of prior investments. We therefore began by creating annual adjacency matrices of all coinvesting activity over the previous five years. For example, the 1990 matrix would aggregate deals done from 1985 to 1989. As in the original, we populated these matrices with all investing entities (those with a unique firm ID code), not just with VC firms. For each replication dataset, we created networks using the corresponding source data and sample.

Fund-Level Results

Table III in HLL2007 reports linear regression estimates of the effect of centrality on VC fund performance. The authors constructed a cross-sectional sample of investment funds, selecting all funds based in the United States, founded between 1980 and 1999, and classified as a VC fund (rather than as, for example, a fund for leveraged buyouts). As a dependent variable, they used the proportion of companies in the fund's portfolio that either were acquired or went public (i.e., had an IPO)—the outcomes associated with the highest returns—by November 2003. To assess the importance of network position, these regressions included the centrality measures for the VC firm that raised the fund, from the year prior to the fund's establishment. Our replication effort followed all of these choices exactly. We also constructed and included all of the control variables used in the original article, according to their descriptions there.

Table 1 presents variable means for HLL2007 and for our three replication datasets. The change in the total number of eligible observations from 3,469 in HLL2007 to 3,464 in our REP1 (2004 download) may reflect some (minor) churn in the underlying database. A slightly larger difference appears in the number of observations in our REP2 (2013 download) data; in REP3, it becomes apparent that over 2,000 VC funds entered the population between 2000 and 2008. Even so, the control variables have highly similar distributions in all datasets, with notable exceptions on fund size, fund inflows, and cumulative parent investments in REP3.

We see greater divergence between the original and the replications in the centrality measures. These differences could have emerged from a variety of factors: (1) difficulties (and different choices) in identifying lead VCs for the directional measures, (2) changes in historical records that affected the network structure, or (3) scaling factors that influenced absolute values but not the distribution of betweenness and eigenvector centrality. None of these factors, however, could entirely account for the magnitude of the differences in the nondirectional degree centrality. In probing further, we discovered that the network variables had almost identical distributions to HLL2007 if the original measures had not been lagged—in other words, if the centrality measures for a 1980 fund, for example, depended on the 1976–1980 network rather than the 1975–1979 network. However, as Hochberg et al. (2007) explicitly state that their analysis used lagged values, we only report regressions using the lagged centrality measures. Note also that we coded eigenvector centrality as zero for funds with isolated VC parent firms (those with no partners at t), roughly 50% of cases in the replication samples.

Table 1: Fund-level variable means for HLL2007 and three replication samples.

Study Label	HLL2007	REP1	REP2	REP3
Study Time Period	1980–1999	1980–1999	1980–1999	1980–2008
Data Download	Early 2004	9/2004	10/2013	10/2013
Exit rate	0.342	0.355	0.340	0.412
IPO rate	0.207	0.226	0.205	0.164
M&A rate	0.136	0.160	0.156	0.263
Fund size	64.000	62.932	60.062	99.646
First fund (0/1)	0.251	0.250	0.252	0.258
Seed or early-stage fund (0/1)	0.365	0.367	0.394	0.466
Corporate VC (0/1)	0.159	0.154	0.147	0.111
VC inflows in vintage year	23.842	31.180	31.575	79.848
Avg B/M ratio in fund's first 3 years	0.514	0.416	0.410	0.384
Aggregate amount parent has invested	71.700	69.563	71.384	180.120
Outdegree centrality	1.203	0.839	0.824	0.691
Indegree centrality	1.003	0.550	0.540	0.442
Degree centrality	4.237	3.353	3.154	2.542
Betweenness centrality	0.285	0.142	0.130	0.093
Eigenvector centrality	3.742	2.024	1.986	1.858
<i>N</i>	3,469	3,464	3,478	5,691

Notes: Statistics for HLL come from HLL2007 Table I. The dependent variables, exit rate, IPO rate, and M&A rate include events that occurred within four years (inclusive) of the end of the study period. The centrality measures report the position of the fund's parent VC firm in the year prior to the fund's vintage year.

The top panel in Table 2 presents our three estimates for fund performance, alongside the corresponding results from Hochberg et al. (2007). Although we had some minor deviations in effect sizes and in statistical significance, the general pattern of results in our replication ended up being very consistent results with the results reported in HLL2007.

Investment-Level Results

Table V in Hochberg et al. (2007) presents probit estimates of the effect of the lead VC firm's centrality in a particular round on the survival of the company receiving that funding. According to the header for their table, the sample in this cross-section included "...portfolio companies that received their first institutional round of funding from a sample VC fund between 1980 and 1999 (and for which relevant cross-sectional information is available)." The dependent variable, survival, equals one if the company received a subsequent funding round or achieved liquidity by November 2003. All other variables had similar definitions as in the fund analysis but stemmed only from the values associated with the lead VC investor. For the sake of simplicity, we only report the results for survival to the second round of funding (HLL2007 also reported survival to the third and fourth rounds). Our

Table 2: VC fund performance and company survival.

VC fund performance	HLL2007	REP1	REP2	REP3
Fund size (ln)	0.029* (0.011)	0.022* (0.006)	0.014* (0.006)	0.016* (0.006)
Fund size (ln) ²	-0.003* (0.002)	-0.002* (0.001)	-0.001 (0.001)	-0.001 (0.001)
First fund	0.010 (0.011)	-0.006 (0.012)	0.014 (0.012)	-0.004 (0.011)
Seed or early-stage fund	-0.023* (0.010)	-0.017 (0.010)	-0.005 (0.010)	-0.020* (0.008)
Corporate VC	0.028 (0.019)	0.025 (0.019)	0.022 (0.019)	0.021 (0.017)
VC inflows (ln)	-0.103* (0.009)	-0.063* (0.007)	-0.068* (0.007)	-0.041 (0.006)
Average B/M ratio	-0.301* (0.030)	-0.263* (0.031)	-0.351* (0.043)	-0.262* (0.029)
Parent experience (ln)	0.008* (0.002)	0.003 (0.004)	0.008* (0.004)	0.009* (0.003)
Eigenvector centrality	0.005* (0.001)	0.010* (0.002)	0.008* (0.002)	0.007* (0.002)
Fund vintage year FE	Yes	Yes	Yes	Yes
<i>N</i>	3,105	3,103	3,107	5,012
Adjusted <i>R</i> ²	0.191	0.193	0.195	0.155
<hr/>				
Portfolio company survival	HLL2007	REP1	REP2	REP3
Fund size (ln)	0.201* (0.030)	0.022* (0.003)	0.018* (0.002)	0.012* (0.001)
Fund size (ln) ²	-0.007 (0.004)	3.33e-4* (4.25e-5)	-2.83e-4* (3.14e-5)	-1.29e-4* (1.48e-5)
First fund	-0.008 (0.026)	0.087* (0.026)	0.056* (0.024)	0.098* (0.015)
Corporate VC	-0.174* (0.055)	-0.099 (0.064)	-0.143* (0.060)	-0.188* (0.034)
VC inflows (ln)	-0.134* (0.021)	-0.168* (0.010)	-0.162* (0.011)	-0.167* (0.006)
Average B/M ratio	-0.563* (0.115)	-0.024 (0.080)	-0.320* (0.116)	0.031 (0.046)
Parent experience (ln)	-0.027* (0.010)	0.073* (0.008)	0.073* (0.008)	0.053* (0.004)
Eigenvector centrality	0.027* (0.003)	0.031* (0.005)	0.029* (0.005)	0.034* (0.004)
Company industry FE	Yes	Yes	Yes	Yes
<i>N</i>	13,761	15,425	17,130	46,984
Pseudo <i>R</i> ²	0.097	0.099	0.099	0.084

Note: Robust standard errors reported in parenthesis; * $p < 0.05$.

analyses of survival to later rounds yielded results consistent with those reported for second round survival.

Although Hochberg et al. (2007) do not report descriptive statistics for the portfolio company dataset, one can infer from HLL2007 Figure 2 that they considered 16,315 companies, with a survival rate of 66.78%, eligible for inclusion in this analysis. Because their probit estimates included 13,761 cases, they apparently dropped 2,554 observations, probably due to missing data. Table 3 reports variable means and sample sizes for each of our portfolio company-level replications. In contrast to the fund replications, our company-level data appear quite different from HLL2007. In REP1 and REP2, we have many more companies, at 26,196 and 30,910, respectively. As we report in Table 2, this difference in sample sizes attenuates substantially when we drop cases with missing data. Arriving at this greater parity, however, required us to relax the restriction—that Hochberg et al. (2007) report—that the lead investor appeared in the fund sample.

The bottom panel in Table 2 presents our three estimates for company survival, alongside the corresponding results from HLL2007. On control variables, we found some important differences from HLL2007 in our replication efforts. In particular, the coefficient on fund size dropped by an order of magnitude, the first fund dummy variable increased by 7–12 times and became statistically significant, and the Book to Market (B/M) ratio coefficients varied across replications. Despite these issues and the sampling mismatches, the coefficient on eigenvector centrality appears quite consistent across all samples.

Data Source Fidelity Over Time

We had the good fortune of having an independent download of the underlying database from a point in time close to that at which the original researchers assembled their data. The fact that we could replicate nearly all of the HLL2007 results, without consulting the authors, is a testament to the completeness of their exposition and, in our opinion, an indicator that they made reasonable judgment calls on variable construction. Tables 1 and 3 nevertheless clearly reveal that our two replications for the 1980–1999 time period draw on substantially different samples.

Commercial databases, such as those sold by Thomson Reuters, generally evolve as the organizations maintaining them collect additional data, change their criteria for the inclusion and exclusion of records, and as they discover and correct errors. We have an opportunity here to inspect and compare the underlying data. Table 4 presents a comparison of key elements for the 1980–1999 investment round data for the 2004 and 2013 downloads. The sample changes on a number of dimensions. For instance, between 2004 and 2013, Thomson added another 5,251 unique portfolio companies to the data (17.8% of the total in the combined sample). Over the same period, it dropped another 486 companies (1.8% of the companies from the 2004 download). These additions and deletions, moreover, do not stem from name changes, as the database includes numerical identification codes (for companies, firms, and funds).

Over this nine-year span, records for events happening more than four years earlier have turnover rates ranging from 6% to 42%, meaning that two scholars

Table 3: Company-level variable means for HLL2007 and three replication samples.

Study Label	HLL2007	REP1	REP2	REP3
Study Time Period	1980–1999	1980–1999	1980–1999	1980–2008
Data Download	Early 2004	9/2004	10/2013	10/2013
Survival (0/1)	0.668	0.596	0.567	0.506
Fund size		202.36	263.461	384.316
First fund (0/1)		0.314	0.315	0.280
Seed or early-stage fund (0/1)		0.269	0.278	0.293
Corporate VC (0/1)		0.038	0.037	0.046
VC inflows in vintage year		39.056	41.466	111.697
Avg B/M ratio in fund's first 3 years		0.478	0.433	0.480
Aggregate amount parent has invested		1275.533	1181.158	2756.944
Outdegree centrality		1.477	3.750	2.509
Indegree centrality		0.974	2.354	1.769
Degree centrality		5.397	9.613	6.760
Betweenness centrality		0.807	2.906	2.146
Eigenvector centrality		2.897	4.085	3.102
<i>N</i>	16,315	26,196	30,910	85,274

Notes: HLL2007 does not report summary statistics for its portfolio company sample; the exit rate in the first column comes from HLL2007 figure 2, whereas the number of observations comes from HLL2007 Table V. The dependent variable, survival, follows companies for four years (inclusive) of the end of the study period and equals one if the portfolio company receives another round of funding, gets acquired, or has an IPO. The VC control variables come from the values associated with the lead VC firm.

studying the same sampling frame at different points in time might analyze quite different data. The largest changes in the data appear in terms of the inclusion of foreign companies and the proportion of companies with an exit via a merger or acquisition (M&A). The dramatic expansion in the number of foreign firms may reflect Thomson Reuters' efforts to expand the geographic scope of the database. Many of the M&A deals appeared between November 2003 and 2013, and therefore one would naturally expect the database to include more of these exits. More surprisingly, events that would seem easily observed and immutable also changed: 515 (15.1%) of the IPOs that had been reported as occurring before November 2003 in the 2004 download disappeared by the 2013 download. We would also note that because the investment network stems from coinvestment, the high rate of turnover in the Company-Round-Firm triads would allow for large differences to emerge in the network structures between the two downloads.

Given that the bulk of the changes here come from the addition of non-U.S. companies—and therefore fall outside the scope of HLL2007 and of our replication—we would not necessarily expect these changes to threaten our conclusions. Even so, we do not regard this issue as a mere curiosity. The sheer volume of turnover, combined with the number of items dropped many years after events presumably occurred, suggests that this churn does not simply reflect random noise or minor error corrections. It nevertheless remains to be seen whether some systematic

Table 4: Comparison of September 2004 and October 2013 downloads for 1980–1999 investment rounds.

	<i>N</i> in 2004 Download	Dropped by 2013	Added by 2013	Turnover rate
Unique Companies	26,960	486	5,251	17.81%
US-based Companies	20,597	451	908	6.32%
Foreign or Unknown Companies	6,363	147	4,455	42.54%
IPO Companies (at 2004 download)	3,454	501	356	22.49%
IPO Companies (through Nov. 2003)	3,400	515	133	18.34%
M&A Companies (at 2004 download)	2,982	244	1762	42.28%
M&A Companies (through Nov. 2003)	2,861	260	467	21.84%
Unique Funds	7,952	460	1,362	19.56%
Unique Firms	4,395	199	733	18.17%
Unique Company-Round-Firm Triads	134,502	21,985	33,438	33.00%

Notes: This table compares unique data points in the disbursement round tables, merging by the appropriate identifier code (company, firm, fund, or round). Turnover rate = $(N \text{ only in 2004} + N \text{ only in 2013})/N$ in both 2004 and 2013. This table focuses on the additions and deletions of entities, but modifications of entity characteristics can also influence some of these counts. For instance, some portfolio companies present in both downloads report different locations (U.S. to non-U.S. and the reverse), which explains why the sum of dropped U.S. and foreign companies in 2013 exceeds the sum of dropped companies in the top row.

pattern underlies these changes and whether they would have a substantial effect on other studies.

Effect Stability

Having established the replicability of the original results and that the use of a more recent download of the data source would not lead to differences in the estimates as an artifact of changes to the database, we then turned to the question of whether the results would vary depending on the sampling frame in terms of the time period being analyzed.

In exploring the importance of the period of study, we focused on aggregate VC investing activity. Boom and bust periods differ in at least one important respect. During booms, one tends to see both a rapid expansion in the number and size of VC funds. Arguably, the supply of financial capital expands far more rapidly during these periods than does the pool of high-quality start-ups in need of funding. Interest during the boom periods also frequently concentrates on a small number of sectors (Gompers et al. 2008; Sorenson and Stuart 2008). As a result, investors end up competing intensely over the most sought-after entrepreneurs and their start-ups. By contrast, during the bust periods, even relatively promising start-ups might find themselves with only a single VC firm interested in investing in them.

Table 5 reports the coefficients of interest for various regressions with alternative dependent variables and sampling frames. Note that each of these models includes, but does not report, all of the control variables included in the models in Table

Table 5: Coefficients for eigenvector centrality on fund performance and company survival by time period.

VC fund performance	Eigenvector centrality	<i>N</i>
Full Sample: 1980–2008	0.007* (0.002)	5,012
Early Period: 1980–1990	0.009* (0.003)	1,652
Industry Shakeout: 1991–1994	0.000 (0.009)	317
Industry Boom: 1995–2001	0.013* (0.003)	1,939
Post-Boom: 2002–2008	–0.002 (0.008)	1,104

Portfolio company survival	Eigenvector centrality	<i>N</i>
Full Sample: 1980–2008	0.034* (0.004)	46,984
Early Period: 1980–1990	0.004 (0.009)	5,121
Industry Shakeout: 1991–1994	0.008 (0.013)	2,358
Industry Boom: 1995–2001	0.039* (0.006)	18,714
Post-Boom: 2002–2008	0.029* (0.008)	20,791

Notes: This table reports results for eigenvector centrality only, using the same model specifications as in Table 2 and using the REP3 (2013 download) dataset. We use VC fund founding dates and portfolio company funding dates to split the samples into periods.

Robust standard errors reported in parenthesis; * $p < 0.05$

2. In the upper panel, for VC fund performance, we first report baseline pooled results for 1980–2008 from Table 2 and then split the sample into distinct time periods based on industry conditions: 1980–1990 (an early boom/growth period), 1991–1994 (a period of low activity), 1995–2001 (the Internet boom), and 2002–2008 (the post-boom period of low activity). In the lower panel, for portfolio company survival, we again report only the coefficient of interest and begin with baseline results for 1980–2008 before splitting the sample into time periods.

The results in Table 5 suggest that the relationship between centrality and performance varies as a function of the period studied. At the fund level, the coefficient for VC centrality only appears significantly different from zero during

boom periods. In fact, the point estimates sometimes even turn negative during cold investing periods. Two different researchers analyzing this setting using identical methods but at different points in time or using different observation windows could come to differing conclusions.

At the company level, the results appear more mixed, with the largest coefficients being in the more recent periods rather than corresponding to the economic cycle. Although these company-level results might appear inconsistent with the fund-level results, note that they examine quite different outcomes. Survival to the next round of funding often represents a necessary condition for eventually having a successful exit through a public offering (IPO) or sale of the company (M&A), but it by no means qualifies as a sufficient one. Start-ups sometimes go through multiple rounds of funding and become cash flow positive but not successful enough to provide an exit for the investor (Ruhnka, Feldman, and Dean 1992). Venture capitalists often refer to these companies as the *living dead*, and they contribute little to the positive performance of a fund. In other cases, the need for additional rounds of investment can even reflect poor performance, as continued negative cash flows require additional infusions (colloquially referred to as “throwing good money after bad”). Park and Steensma (2011), for example, reported a negative relationship between the number of funding rounds and the probability of going public.

The fact that the value of centrality varies across points in the economic cycle may also have theoretical implications. At least four factors might generate a positive relationship between centrality and performance: (1) More extensive networks give venture capitalists better access to private information about entrepreneurs and their companies, thereby allowing them to choose among a broader range of investments with more extensive—and perhaps higher quality—information on each (Sorenson and Stuart 2001). (2) Those in central positions also have better information about the supply of and demand for VC and therefore a better sense of the appropriate prices for these private transactions (Hochberg et al. 2007). (3) Centrality, to the extent that it engenders and reflects status within the industry, should also provide VC firms with priority in their ability to access sought-after investment syndicates (Podolny 1993; Sorenson and Stuart 2001). Kleiner Perkins has much better odds of being invited to invest in the latest hot deal, a Facebook or Google, than a no-name venture investor. (4) Finally, centrality may aid the nurturing of portfolio companies, to the extent that the venture capitalist has more extensive connections among buyers, suppliers, and potential employees that it can share with its portfolio companies.

Of these, the third—the ability to invest in popular portfolio companies—seems the one most sensitive to economic conditions. High-quality due diligence and the ability to price deals correctly always matter and therefore should prove valuable regardless of the period. Mentoring and monitoring, moreover, may actually become *less* important in a climate in which public investors have such an appetite for equity in an emerging industry that they pay limited attention to profitability (Gompers et al. 2008).

But the benefits of status emerge primarily in the context of competition for deals. When an entrepreneur only receives a single offer of investment, she usually has little choice but to accept it. By contrast, when entrepreneurs can choose among

investors and venture capitalists themselves have more potential syndicate partners than they can accommodate, high-status venture capitalists usually receive a place at the front of these queues (Podolny 1993; Hsu 2004). During busts and even during normal economic periods, start-ups rarely have investors competing to give them money. Only during the waves of excess liquidity do the valuations become frothy and the competition to invest in certain companies intense.

Discussion

Hochberg et al. (2007) provided one of the first studies to connect centrality in the interorganizational network to the financial performance of firms. It remains, to date, one of the best studies on this topic. But would Hochberg et al. (or someone else) have arrived at the same conclusions if they had done their study at a somewhat different time, a few years earlier or later? Two issues come into play. First, the underlying database maintained by Thomson Reuters changes over time, even for data that one might expect to remain constant. Second, the results may depend sensitively on the environmental conditions present during the time of the sampling frame.

We found that the Hochberg et al. (2007) results remained largely robust to the first issue. Although Thomson Reuters retroactively changed data at a high rate—altering nearly half of the observations within certain subsets—these changes did not appear to influence the HLL2007 results. Even in this near-exact replication, however, some small differences emerged in the sample sizes and variable distributions. These differences point to an issue with ThomsonONE (formerly, Venture Economics and then VentureXpert) that has not previously received attention: the tendency for its historical records to change over time.

By contrast, HLL2007 does appear somewhat a product of its time. When we examined a longer range of data and in shorter subsets, centrality often had no relationship—and sometimes even had a negative one—with firm performance. The positive relationship appeared most strongly during the boom period of the late 1990s. Because these investments accounted for a large share of the population analyzed by HLL2007, they largely drove the average results in their article.

Attempts at replication in sociology and in nonexperimental social science research more broadly have been limited and usually have remained private—perhaps as an exercise in a research methods course. To the extent that the primary issues in replication involve checking to see whether the researcher made any coding or estimation errors, this relegation of replication to the background may even seem appropriate to many. But archival research depends also on (usually unstated) assumptions about the stability of the data: (1) that the archive faithfully reflects the characteristics of the population (but, to the extent that it does not, its deviation from the population does not change over time) and (2) that the population and processes do not change in their distribution and dynamics over time. The first of these issues poses a difficult challenge; addressing it may require researchers with snapshots of a database from different points in time to join forces. Researchers, however, can address the second one more readily. By examining temporal subsets

of the data, they can examine the extent to which the relationships observed seem relatively immutable versus being subject to some important scope conditions.

References

- Bonacich, Phillip. 1987 "Power and Centrality – a Family of Measures." *American Journal of Sociology* 92:1170–1182. <https://doi.org/10.1086/228631>
- Breznau, Nate. 2015. "The Missing Main Effect of Welfare State Regimes: A Replication of 'Social Policy Responsiveness in Developed Democracies' by Brooks and Manza" *Sociological Science* 2:420–441. <https://doi.org/10.15195/v2.a20>
- Gompers, Paul, Anna Kovner, Josh Lerner, and David Sharfstein. 2008 "Venture Capital Investment Cycles: The Impact of Public Markets." *Journal of Financial Economics* 87:1–23. <https://doi.org/10.1016/j.jfineco.2006.12.002>
- Hannon, Lance, and Robert Defina. 2016. "Can Incarceration Really Strip People of Racial Privilege?" *Sociological Science* 3:190–201. <https://doi.org/10.15195/v3.a10>
- Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu. 2007 "Whom You Know Matters: Venture Capital Networks and Investment Performance." *Journal of Finance* 62:251–301. <https://doi.org/10.1111/j.1540-6261.2007.01207.x>
- Hsu, David H. "What Do Entrepreneurs Pay for Venture Capital Affiliation?" *Journal of Finance* 59:1805–1844. <https://doi.org/10.1111/j.1540-6261.2004.00680.x>
- Ioannidis, John P. A. 2005 "Why Most Published Research Findings Are False." *PLoS Medicine* 2:e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Lee, Byungkyu, and Peter Bearman. 2017. "Important Matters in Political Context" *Sociological Science* 4:1–30. <https://doi.org/10.15195/v4.a1>
- Ljungqvist, Alex, Christopher Malloy, and Felicia Marston. 2009. "Rewriting History." *Journal of Finance* 64:1935–60. <https://doi.org/10.1111/j.1540-6261.2009.01484.x>
- Park, Haemin Dennis, and H. Kevin Steensma. 2012 "When Does Corporate Venture Capital Add Value for New Ventures." *Strategic Management Journal* 33:1–22. <https://doi.org/10.1002/smj.937>
- Phalippou, Ludovic, and Oliver Gottschalg. 2009 "The Performance of Private Equity Firms." *Review of Financial Studies* 22:1747–1776. <https://doi.org/10.1093/rfs/hhn014>
- Podolny, Joel M. 1993 "A Status-Based Model of Market Competition." *American Journal of Sociology* 98:829–872. <https://doi.org/10.1086/230091>
- Podolny, Joel M. 2001 "Networks as Pipes and Prisms of the Market." *American Journal of Sociology* 107:33–60. <https://doi.org/10.1086/323038>
- Ruhnka, John C., Howard D. Feldman, and Thomas J. Dean. 1992 "The 'Living Dead' Phenomenon in Venture Capital Investments." *Journal of Business Venturing* 7:137–155. [https://doi.org/10.1016/0883-9026\(92\)90009-G](https://doi.org/10.1016/0883-9026(92)90009-G)
- Samila, Sampsa, and Olav Sorenson. 2011. "Venture Capital, Entrepreneurship, and Economic Growth." *Review of Economics and Statistics* 93:338–349. https://doi.org/10.1162/REST_a_00066
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011 "False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22:1359–1366. <https://doi.org/10.1177/0956797611417632>

- Sorenson, Olav, and Toby E. Stuart. 2001. "Syndication Networks and the Spatial Distribution of Venture Capital Investments." *American Journal of Sociology* 106:1546–1588. <https://doi.org/10.1086/321301>
- Sorenson, Olav, and Toby E. Stuart. 2008. "Bringing the Context Back In: Settings and the Search for Syndicate Partners in Venture Capital Investment Networks." *Administrative Science Quarterly Journal* 53:266–294. <https://doi.org/10.2189/asqu.53.2.266>
- Sørensen, Morten. 2007. "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital" *Journal of Finance* 62:2725–2762. <https://doi.org/10.1111/j.1540-6261.2007.01291.x>
- Trapido, Denis. 2007. "Competitive Embeddedness and the Emergence of Interfirm Cooperation." *Social Forces* 86:165–191. <https://doi.org/10.1353/sof.2007.0110>

Yuan Shi: Robert H. Smith School of Business, University of Maryland.
E-mail: yuanshi@rhsmith.umd.edu.

Olav Sorenson: Yale School of Management, Yale University.
E-mail: olav.sorenson@yale.edu.

David M. Waguespack: Robert H. Smith School of Business, University of Maryland.
E-mail: dwaguesp@rhsmith.umd.edu.