

# Internet Appendix: “Are Early Stage Investors Biased Against Women?”

Michael Ewens<sup>a</sup>, Richard R. Townsend<sup>b,\*</sup>

<sup>a</sup>*California Institute of Technology, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125*

<sup>b</sup>*University of California San Diego, Rady School of Management, 9500 Gilman Dr, La Jolla, CA 92093*

---

## Abstract

This document provides supplementary material for the paper "Are Early Stage Investors Biased Against Women?" Section A discusses how our startup level analysis relates to startup-investor pair level analysis. Section B provides supplementary tables.

---

---

\*Corresponding author. Tel: 858-246-1632; Fax: 858-534-0745;  
Email address: [rرتownsend@ucsd.edu](mailto:rرتownsend@ucsd.edu) (Richard R. Townsend).

## A. Startup level versus Startup-Investor pair level analysis

Our baseline analysis is at the startup level. In particular, we estimate equations of the form:

$$YM_i = \alpha + B_m FEnt_i + \boldsymbol{\delta}' \mathbf{X}_i + \epsilon_i, \quad (\text{A1})$$

where  $i$  indexes startups,  $YM_i$  represents the number of male investors that expressed interest in startup  $i$ ,  $FEnt_i$  is an indicator equal to one if startup  $i$  is led by a female entrepreneur, and  $\mathbf{X}_i$  is a vector of startup/entrepreneur level controls.<sup>1</sup> For comparison, we also estimate:

$$YF_i = \alpha + B_f FEnt_i + \boldsymbol{\delta}' \mathbf{X}_i + \epsilon_i, \quad (\text{A2})$$

where in this case  $YF_i$  represents expressions of interest by female investors in startup  $i$ .

Alternatively, one could do the same analysis at the startup-investor pair level, estimating equations of the form:

$$y_{ij} = \alpha + b_m FEnt_i + \boldsymbol{\delta}' \mathbf{X}_i + \epsilon_{ij}, \quad \forall j \text{ where gender}(j)=\text{male} \quad (\text{A3})$$

$$y_{ij} = \alpha + b_f FEnt_i + \boldsymbol{\delta}' \mathbf{X}_i + \epsilon_{ij}, \quad \forall j \text{ where gender}(j)=\text{female} \quad (\text{A4})$$

where  $j$  indexes investors, and  $y_{ij}$  is an indicator equal to one if investor  $j$  expressed interest in startup  $i$ . In other words, each startup observation is matched with each investor and Equation A3 is estimated using pairs involving male investors, while Equation A4 is estimated using pairs involving female investors. Let,  $I_m$  and  $I_f$  represent the number of male and female investors on AngelList, respectively. Then it must be the case the  $b_m = B_m/I_m$  and  $b_f = B_f/I_f$ . Thus, the pairwise analysis simply re-scales the coefficients. The re-scaling is somewhat arbitrary in the sense

---

<sup>1</sup>In practice, rather than using count variables for our outcomes, we use indicator variables equal to one if the startup receives interest from any number of investors. We do this minimize the influence of outliers. It also separates the extensive margin from the intensive margin. However, in our data, most startups receive interest from zero or one investors in any case. We show a count variable version of our results in Appendix Table A2.

that it depends on how one defines the pool of potential investors. For example, it will depend on whether investors who never express interest in any startup are included.<sup>2</sup>

Another alternative would be to pool male and female investors in the pairwise analysis and estimate a single equation:

$$y_{ij} = \alpha + \beta_1 FEnt_i \times FInv_j + \gamma FEnt_i + \theta FInv_j + \delta'_1 \mathbf{X}_i + \delta'_2 (\mathbf{X}_i \times FInv_j) + \epsilon_{ij}. \quad (\text{A5})$$

Here  $FInv_j$  is an indicator equal to one if investor  $j$  is a woman. In this case,  $\beta_1 = b_f - b_m = B_f/I_f - B_m/I_m$ . In addition to not providing any additional information, focusing on  $\beta_1$  as a measure of male investor bias requires the assumption that the behavior of female investors is unbiased. We want to remain open to the possibility that female investors are biased in favor of female entrepreneurs. One could also estimate Equation A5 with startup fixed effects,  $\eta_i$ :

$$y_{ij} = \alpha + \beta_2 FEnt_i \times FInv_j + \theta FInv_j + \delta'_2 (\mathbf{X}_i \times FInv_j) + \eta_i + \epsilon_{ij}. \quad (\text{A6})$$

However, because each startup faces the same set of potential investors, the estimated coefficient on the key interaction term will remain exactly the same:  $\beta_2 = \beta_1$ . Similarly, one could also estimate Equation A5 with both startup fixed effects,  $\eta_i$ , and investor fixed effects,  $\eta_j$ :

$$y_{ij} = \alpha + \beta_3 FEnt_i \times FInv_j + \delta'_2 (\mathbf{X}_i \times FInv_j) + \eta_i + \eta_j + \epsilon_{ij}. \quad (\text{A7})$$

Again, the estimated coefficient on the key interaction term will remain exactly the same:  $\beta_3 = \beta_1$ .

An additional potential benefit of pairwise analysis is the possibility of including pairwise control variables. For example, in the pairwise regression, one could include an indicator variable equal to

---

<sup>2</sup>Relatedly, one could refine Equations A3 and A4, by excluding from the sample startup-investor pairs involving investors who had not yet joined AngelList yet as of the date of the startup's fundraising campaign. Doing so also requires some assumptions, as investors may join AngelList after the start of fundraising campaign but still express interest in a company. More generally, the issue of changing investor composition is controlled for in both the startup-level and pair-level analysis by including fundraising year fixed effects in the vector of controls,  $\mathbf{X}_i$ .

one if the entrepreneur and investor are located in the same region. In principle, we could still control for this at the startup level by controlling for the percent of investors located in the same region as startup  $i$ . However, a more general way of controlling for regional variation at the startup level is simply to include startup region fixed effects. These fixed effects would absorb the variable representing the percent of investors located in the same region and also allow for regional variation that is not related to investor proximity. Overall, no information is added by going to the startup-investor pair level since our key variable of interest, entrepreneur gender, only varies at the startup level.

## B. Supplementary Tables

**Table A1**

Variable definitions

This table describes the main variables used throughout the analysis.

Variable	Definition
Shared	An indicator equal to one if the startup was shared by a male (female) investor on the AngelList platform.
Received introduction	An indicator equal to one if the startup received at least one introduction request from a male (female) investor on the AngelList platform. An introduction is a one-way inquiry by an investor to a startup firm. Introduction requests are only possible when the startup has an active fundraising campaign. Direct communication between investor and startup is not possible on the platform without an accepted introduction request.
Funded	An indicator equal to one if the startup was observed as raising capital after the start of their fundraising campaign from an investor on AngelList.
Startup failed	An indicator equal to one if the startup if the startup's website was no longer active as of November 2016. We deem a website as inactive if it fails to load and/or if its domain is available for purchase.
Had IPO or Acquisition	An indicator equal to one if the startup firm had an initial public offering or was acquired by November 2016. Such exits are observed in either Crunchbase or VentureSource.
Team size	The number of founders of the startup as listed in the startup's AngelList profile.
Year fundraised	The year the startup's fundraising campaign first began on AngelList.
Capital sought	The original capital amount sought in the startup's first fundraising campaign on AngelList.
Previous founder	An indicator equal to one if the startup founder's LinkedIn profile indicated a past title as "founder" or "co-founder" at a another firm prior to the founding of the current startup.
Bachelors degree	An indicator equal to one if the startup founder's LinkedIn profile indicated that they received a bachelor's degree or equivalent. Such information is also found in the short founder biography on their AngelList profile.
MD / PhD / JD and MBA	An indicator equal to one if the founder had a PhD, MD or JD (MBA) in their AngelList or LinkedIn profile.
Elite school (any)	An indicator equal to one if any of the founder's pre-startup degrees were from any of the following universities: MIT, Princeton, UPenn, U. Chicago, Harvard, Yale, Caltech, John Hopkins, Duke, Stanford, Yale, Columbia or Northwestern.
Years experience pre-startup	A count of the number of years from the first observed job date to the founding of the startup as available on the founder's LinkedIn profile.
Age	The age in years of a startup founder based on their year of college graduation, assuming age 22 at graduation.
Attended incubator	A dummy variable equal to one if the startup attended an incubator prior to fundraising.

**Table A2**

Investor interest and funding measured as counts

This table repeats the analysis in Tables 4–6 where we now consider the raw count of the number of investors signaling interest in the startup for each interest type. The columns “Male” consider the number of shares, introductions and funders that are male, while the “Female” columns represent the same but for female investors. The number of introductions is winsorized at the 5% level. All specifications are negative binomial regressions, but the results are qualitatively similar for OLS. All controls and fixed effects are as defined in Table 4. All fixed effects from Tables 4–6 are included in the regressions. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Number shares		Number intros.		Number funders	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
Female	-0.81*** (0.14)	2.08*** (0.10)	-0.18*** (0.067)	0.49*** (0.12)	-0.67*** (0.24)	0.94*** (0.29)
Previous founder	0.37*** (0.066)	0.41*** (0.12)	0.22*** (0.057)	0.49*** (0.11)	0.25 (0.16)	0.63** (0.27)
Bach. degree	0.20*** (0.062)	0.23* (0.13)	0.20*** (0.052)	0.22** (0.11)	0.13 (0.15)	-0.37 (0.25)
PhD/MD/JD	-0.27** (0.12)	0.026 (0.21)	0.098 (0.13)	-0.093 (0.21)	0.059 (0.26)	0.54 (0.68)
MBA	-0.072 (0.085)	-0.39** (0.16)	-0.062 (0.070)	-0.10 (0.14)	0.044 (0.22)	0.27 (0.37)
Elite school (any)	0.25** (0.099)	-0.20 (0.19)	0.29*** (0.072)	0.17 (0.13)	0.15 (0.20)	-0.025 (0.49)
Has traction	0.42*** (0.080)	0.55*** (0.14)	0.45*** (0.061)	0.41*** (0.15)	0.27 (0.17)	0.050 (0.27)
Attended incubator	0.98*** (0.076)	0.98*** (0.13)	1.08*** (0.053)	1.39*** (0.099)	1.57*** (0.14)	1.77*** (0.23)
Log capital sought	0.12*** (0.021)	0.014 (0.044)	0.20*** (0.021)	0.31*** (0.043)	0.33*** (0.056)	0.17 (0.11)
Observations	17780	17780	17780	17780	17780	17780
Pseudo- $R^2$	0.24	0.26	0.13	0.19	0.13	0.16
Model FE?	Neg bin. Y	Neg bin. Y	Neg bin. Y	Neg bin. Y	Neg bin. Y	Neg bin. Y

**Table A3**

## Alternative funding variable

This table reports the linear probability model where the dependent variable is one if the founder was observed successfully raising their first observed round of financing on AngelList, VentureSource or Crunchbase from a male investor by the end of the sample period (11/2015). We supplement AngelList's investment data with three additional sources. First, we match our sample with startups in Dow Jones' VentureSource database. This allows us to identify companies in our sample that eventually raised money from VCs. Second, we match our sample to startups that report raising capital on Crunchbase. Crunchbase's coverage is likely to be better than VentureSource for seed rounds with no institutional investor. Third, we found startups on Pitchbook and identified the lead investors in the first financing round that occurred after the AngelList fundraising campaign. Unfortunately, our data sources often fail to identify the individual investors involved in a round. This either happens because no investors are identified (only the fact that a financing round occurred is recorded by the data source), or because the investors identified are institutions rather than individuals. For institutional financing rounds from VentureSource, we are able determine the gender of the individual investor who sourced the deal using board membership. That is, we assume the individuals who took board seats in the first financing round were the ones who sourced the deal. In the cases where Crunchbase and Pitchbook have individual investors assigned to specific financing rounds (e.g. as lead partners or angels), we assign gender using those individuals associated with the first financing that occurs after the fundraising campaign on AngelList. A unit of observation is a US-based startup on the platform where we can identify the gender of all the founders and where the capital sought is at least \$5000. The funding dependent variable is one if the startup raised capital from an investor identified as male and listed on AngelList, VentureSource or Crunchbase. Variables are as defined in Table A1. All the "FE" are as defined in Table 4. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Startup was funded by a					
	Male investor			Female investor		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.017*** (0.0028)	-0.012*** (0.0028)	-0.012*** (0.0028)	0.0068*** (0.0023)	0.0079*** (0.0022)	0.0079*** (0.0023)
Previous founder			0.016*** (0.0048)			0.0042* (0.0025)
Bach. degree			0.011*** (0.0029)			-0.0011 (0.0015)
PhD/MD/JD			-0.0020 (0.0099)			-0.0056 (0.0043)
MBA			-0.0075 (0.0064)			-0.0015 (0.0031)
Elite school (any)			0.014** (0.0070)			0.0033 (0.0035)
Has traction			0.0039 (0.0061)			-0.00099 (0.0030)
Attended incubator			0.079*** (0.0076)			0.027*** (0.0044)
Log capital sought			0.0049*** (0.00076)			0.0013*** (0.00042)
Observations	17780	17780	17780	17780	17780	17780
$R^2$	0.014	0.055	0.078	0.0038	0.016	0.025
Round year FE?	Y	Y	Y	Y	Y	Y
Firm join year FE?	Y	Y	Y	Y	Y	Y
Team size FE?	N	Y	Y	N	Y	Y
Industry FE?	N	Y	Y	N	Y	Y
Location FE?	N	Y	Y	N	Y	Y

**Table A4**

Solo founders

This table repeats the analysis in Tables 4-6 for the subset of startups with only one founder. All fixed effects from Tables 4–6 are included in the regressions. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Male investors			Female investors		
	Shared (1)	Introduced (2)	Funded (3)	Shared (4)	Introduced (5)	Funded (6)
Female	-0.011*** (0.0026)	-0.021*** (0.0065)	-0.0061*** (0.0018)	0.0048*** (0.0016)	0.010*** (0.0034)	0.0047*** (0.0016)
Previous founder	0.024*** (0.0053)	0.032*** (0.0097)	0.0031 (0.0036)	0.0024 (0.0015)	0.0093** (0.0045)	0.0020 (0.0016)
Bach. degree	0.0084*** (0.0028)	0.033*** (0.0059)	0.0021 (0.0020)	0.00065 (0.00072)	0.0059** (0.0026)	-0.00025 (0.00080)
PhD/MD/JD	0.0061 (0.011)	0.0078 (0.020)	0.010 (0.0087)	0.0014 (0.0033)	0.0078 (0.010)	0.0023 (0.0037)
MBA	0.0054 (0.0072)	0.0063 (0.013)	0.0050 (0.0052)	0.0019 (0.0023)	0.0041 (0.0064)	0.0049* (0.0028)
Elite school (any)	0.027*** (0.0080)	0.045*** (0.013)	0.0075 (0.0055)	0.0022 (0.0024)	0.014** (0.0069)	-0.0020 (0.0017)
Has traction	0.033*** (0.0076)	0.094*** (0.013)	0.014*** (0.0053)	0.0053** (0.0025)	0.013** (0.0060)	-0.0024* (0.0014)
Attended incubator	0.085*** (0.010)	0.18*** (0.015)	0.041*** (0.0075)	0.0049* (0.0028)	0.067*** (0.0093)	0.0099*** (0.0036)
Log capital sought	0.0037*** (0.00074)	0.011*** (0.0017)	0.0018*** (0.00051)	0.00032 (0.00021)	0.0030*** (0.00064)	0.00050** (0.00024)
Observations	14211	14211	14211	14211	14211	14211
$R^2$	0.073	0.11	0.020	0.0078	0.041	0.0055
FEs?	Y	Y	Y	Y	Y	Y



**Table A5**

Excluding social mission investors

This table repeats the analysis found in Table 4-6. The investor interest measures here exclude signals of interest made by female investors on the AngelList platform associated with a fund with a social mission to promote female entrepreneurship. These funds are Valor Ventures, Golden Seeds, Pipeline Angels, Built By Girls Ventures, BELLE Capital USA, Female Founders Fund, The Womens' Venture Capital Fund, Forerunner Ventures, 500 Women, Angel Academe, Phenomenelle Angels Fund, Broadway Angels, Topstone Angels, Plum Alley, The Jump Fund, Astia, Scale, Cross Culture Ventures, Gotham Gal Ventures, True Wealth Ventures, Halogen Ventures, Sofia Fund, Female Funders, Women Angels, Women Founders Network, Women Launch and Women Lead Inc. A unit of observation is a US-based startup on the platform where we can identify the gender of all the founders and where the capital sought is at least \$5000. Variables are as defined in Table A1. All the "FE" are as defined in Table 4. Robust standard errors reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Female investors, excluding social mission investors					
	Shared		Introduced		Funded	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.076*** (0.0053)	0.078*** (0.0053)	0.013*** (0.0042)	0.018*** (0.0042)	0.0026 (0.0017)	0.0034** (0.0017)
Previous founder		0.010** (0.0041)		0.017*** (0.0050)		0.0033 (0.0021)
Bach. degree		0.0059** (0.0025)		0.0089*** (0.0030)		-0.00091 (0.0012)
PhD/MD/JD		0.0081 (0.0085)		-0.0020 (0.010)		-0.0027 (0.0035)
MBA		-0.0042 (0.0051)		-0.0061 (0.0067)		-0.00093 (0.0025)
Elite school (any)		-0.0022 (0.0052)		0.018** (0.0073)		-0.00052 (0.0026)
Has traction		0.021*** (0.0053)		0.019*** (0.0065)		-0.00045 (0.0024)
Attended incubator		0.038*** (0.0057)		0.096*** (0.0079)		0.019*** (0.0035)
Log capital sought		0.00012 (0.00074)		0.0044*** (0.00080)		0.00043 (0.00035)
Observations	17780	17780	17780	17780	17780	17780
$R^2$	0.053	0.076	0.019	0.094	0.0016	0.015
Round year FE?	Y	Y	Y	Y	Y	Y
Firm join year FE?	Y	Y	Y	Y	Y	Y
Team size FE?	N	Y	N	Y	N	Y
Industry FE?	N	Y	N	Y	N	Y
Location FE?	N	Y	N	Y	N	Y

**Table A6**

Excluding startups in consumer industries

This table reports the regressions found in Tables 4-6 after excluding startups in the consumer services and consumer products industries. A unit of observation is a US-based startup on the platform where we can identify the gender of all the founders and where the capital sought is at least \$5000. Variables are as defined in Table A1. All controls and fixed effects from Tables 4-6 are included in the regressions. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Shared by		Intro. from		Funded by	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
Female	-0.024*** (0.0053)	0.0098*** (0.0036)	-0.028** (0.011)	0.016** (0.0073)	-0.014*** (0.0043)	0.00013 (0.0027)
Observations	8618	8618	8618	8618	8618	8618
$R^2$	0.14	0.018	0.20	0.11	0.053	0.014
Controls?	Y	Y	Y	Y	Y	Y
FEs?	Y	Y	Y	Y	Y	Y

**Table A7**

Gender neutral startups based on mean response

This table reports the regressions found in Table 8 after excluding startups where the average survey respondent labeled the firm as gender-neutral. A unit of observation is a US-based startup on the platform where we can identify the gender of all the founders and where the capital sought is at least \$5000. Variables are as defined in Table A1. All controls and fixed effects from Tables 4-6 are included in the regressions. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Shared by		Intro. from		Funded by	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
Female	-0.011*** (0.0039)	0.0086*** (0.0024)	-0.016** (0.0076)	0.020*** (0.0049)	-0.0082*** (0.0029)	0.0048** (0.0022)
Observations	14692	14692	14692	14692	14692	14692
$R^2$	0.13	0.019	0.19	0.099	0.050	0.016
Controls?	Y	Y	Y	Y	Y	Y
FEs?	Y	Y	Y	Y	Y	Y

**Table A8**

Investor heterogeneity: Pigeonholing at tails

This table repeats the analysis of Table 13, replacing the continuous variables *Log Capital Sought* and *Percent Female Industry* with indicator variables *Low Capital Sought* and *High Percent Female Industry*, equal to one when the underlying continuous variables are in the bottom or top quintile of the sample, respectively. A unit of observation is a US-based startup on the platform where we can identify the gender of all the founders and where the capital sought is at least \$5000. Variables are as defined in Table A1. All controls and fixed effects from Tables 4–6 are included in the regressions, except in Panel B where industry fixed effects are excluded. Robust standard errors are reported in parentheses. Significance: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

	Panel A: Low capital sought					
	Male investors			Female investors		
	Shared (1)	Introduced (2)	Funded (3)	Shared (4)	Introduced (5)	Funded (6)
Female	-0.019*** (0.0040)	-0.019** (0.0080)	-0.012*** (0.0029)	0.0096*** (0.0025)	0.020*** (0.0052)	0.0048** (0.0022)
Low capital sought	-0.013*** (0.0028)	-0.030*** (0.0070)	-0.0030 (0.0023)	0.0017** (0.00082)	-0.0030 (0.0026)	0.0016 (0.0012)
Female X Low capital sought	0.024*** (0.0053)	-0.010 (0.012)	0.0095*** (0.0035)	-0.0089*** (0.0025)	-0.013* (0.0065)	-0.0055** (0.0024)
Observations	17780	17780	17780	17780	17780	17780
$R^2$	0.12	0.19	0.049	0.018	0.094	0.015
Controls?	Y	Y	Y	Y	Y	Y
FEs?	Y	Y	Y	Y	Y	Y
	Panel B: High percent female industry					
	Male investors			Female investors		
	Shared (1)	Introduced (2)	Funded (3)	Shared (4)	Introduced (5)	Funded (6)
Female	-0.025*** (0.0053)	-0.034*** (0.011)	-0.011** (0.0042)	0.011*** (0.0038)	0.015** (0.0073)	0.0070* (0.0036)
High Percent Female Industry	-0.0075** (0.0034)	-0.025*** (0.0064)	-0.0013 (0.0027)	0.0011 (0.0010)	-0.0051* (0.0031)	0.00012 (0.0012)
Female X High % female ind.	0.024*** (0.0070)	0.035** (0.014)	0.0028 (0.0052)	-0.0048 (0.0044)	0.0086 (0.0090)	-0.0044 (0.0041)
Observations	17780	17780	17780	17780	17780	17780
$R^2$	0.11	0.16	0.035	0.016	0.071	0.0092
Controls?	Y	Y	Y	Y	Y	Y
FEs?	Y	Y	Y	Y	Y	Y