Economics   Leavey School of Business

4-13-2016

Do Gender-Variant Preferences for Competition Persist in the Absence of Performance?

John Ifcher
Santa Clara University, jifcher@scu.edu

Homa Zarghamee

Do Gender-Variant Preferences for Competition Persist in the Absence of Performance?

John Ifcher and Homa Zarghamee[1]

JEL codes: J16, C91, J24

Keywords: Preference for Competition, Gender Differences, Performance, Willingness to Compete; Task without Agency

---

[1] Ifcher: Department of Economics, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, (408) 554-5579, jifcher@scu.edu. Zarghamee: Department of Economics, Barnard College, 3009 Broadway, New York, NY 10027, (212) 854-8946, hzargham@barnard.edu. The authors contributed equally to this work.

1. Introduction

Gender differences in the willingness to compete are believed to have a profound impact on labor market outcomes and may help explain the paucity of women in senior management positions (Niederle, Segal, & Vesterlund, 2013). Disparities in labor market outcomes persist even in the face of female educational gains and greater government guarantees against gender discrimination (Goldin, Katz, & Kuziemko, 2006). Other factors—such as gender differences in communication and negotiation styles, maternity leave and the consequent beliefs about labor-force attachment, and gender discrimination—surely also help explain occupational segregation and the gender wage gap.

In their seminal paper, Niederle and Vesterlund (NV) (2007) show that, in a real-effort summation task, men are more willing to compete than women; i.e., men are more likely than women to choose to be paid according to a winner-take-all tournament (WTA) as opposed to a piece-rate payment scheme (PR). NV explore various explanations for the gender difference in willingness to compete, specifically, gender-variant overconfidence, risk preference, feedback aversion (aversion to information about performance relative to others), and preference for competition. To be clear, there is an important distinction between the gender difference in ***willingness to compete*** and ***preference for competition***. For example, a gender difference in willingness to compete, if fully explained by overconfidence would suggest no gender difference in the preference for competition.

An important question is whether NV's experiment revealed a novel, heretofore, unidentified preference, the ***preference for competition***, and a gender difference in it; or whether their finding is an artifact of previously known gender differences. If the former is true, NV's contribution is fundamental, deepening our understanding of preferences, gender, and decision-making. Moreover, knowing whether it is a novel preference helps shape policy responses to ameliorate the disparities that arise from gender differences in willingness to compete. NV conclude that there is a gender difference in the preference for competition, and that it is explained by gender differences in the "thrill or fear of ***performing*** in a competitive environment" (italics added).

While NV provide strong evidence that gender-variant overconfidence (not risk preferences or feedback aversion) is an important determinant of the gender difference in willingness to compete, the evidence regarding performing in a competitive environment requires further investigation. In this paper we carefully examine the importance of performing by comparing willingness to compete in tasks with and without agency. Specifically, we create a new task that eliminates performance altogether, which we call the ***task without agency***. We find evidence that performing in a competition is sufficient, but may not be necessary, to observe gender-variant preferences for competition.

This is important for several reasons. First, if performance is sufficient to generate gender-variant preferences for competition, then our research informs the structure of the utility function, with performance as a determinant of the preference for competition. Further, if overconfidence, but not performance, were sufficient, then gender-variant willingness to compete could be eliminated by correcting individuals' beliefs with information about their relative performance. If performance is necessary, then adequately routinizing jobs (i.e., eliminating idiosyncratic performance) should eliminate the gender difference in the preference for competition. If performance is not necessary, then, in the absence of performance, other factors—e.g., noise, social preferences, and preferences over the pay structure itself—must drive the gender gap in the preference for competition. For example, women may be less willing to be hedge fund managers, not only because they fear performing in a competitive environment, but also because the fund's returns, and hence their pay, are determined by forces outside of their control.

2. Background and theoretical framework

NV attempt to identify the gender difference in the ***preference for competition*** by demonstrating that the three other plausible explanations in their theoretical model—overconfidence, risk preferences, and feedback aversion—do not fully explain the gender difference in ***willingness to compete***. The residual gender gap in willingness to compete (controlling for overconfidence, risk preferences, and feedback aversion) is theorized to be the result of variant preferences for competition. Moreover, they equate the preference for competition with the "thrill or fear of ***performing*** in a competitive environment" (italics added). Below, we review NV's empirical approach. Then we present weaknesses in their identification strategy and theoretical model that our paper will attempt to address.

NV's identification strategy unfolds as follows. First, they demonstrate that men and women perform equally well in the summation task, ruling out that differences in actual performance drive the gender gap in willingness to compete. They then control for beliefs about performance in their regression analysis, as the choice to compete is made before performing (ex ante). While willingness to compete increases in beliefs about performance, and men are more overconfident than women, a significant, albeit smaller, residual gender gap remains. From these results, NV conclude that male overconfidence explains a portion of the gender difference in the willingness to compete. Next, NV examine subjects' willingness to compete when the choice to compete is made after the task is completed (ex post), as opposed to ex ante. They argue that the salient difference between ex ante and ex post choices is whether one ***performs*** subsequent to choosing to compete, since risk and feedback are present in both choices. They find that the residual gender gap is insignificant with ex post choices and that a significant residual gender gap remains with ex ante choices controlling for ex post choices. They conclude that the gender-variant preference for

competition, i.e., the residual gender gap with ex ante choices, is the result of neither risk preferences nor feedback aversion, but rather gender-variant feelings about performing in a competitive environment.

NV (2007) inspired a series of laboratory experiments that test the robustness and limits of their finding. Thorough reviews of this research can be found in NV (2011) and Niederle (forthcoming). In NV (2011), the authors organize the studies by the residual-gender-gap explanation tested in the research; two of the explanations are included in the NV (2007) theoretical model (beliefs and risk preferences) and one is new (other-regarding preferences). We organize our discussion of the literature by these residual-gender-gap explanations.

To examine the role of ***beliefs***, many studies have followed the NV (2007) methodology of eliciting and controlling for self-reported rank. The result—that gender-variant beliefs explain some but not all of the gender gap in willingness-to-compete—is replicated in most studies (e.g., Almås et al., 2014; Balafoutas, Kerschbamer & Sutter, 2012; Balafoutas & Sutter, 2012; Buser, Niederle, & Oosterbeek, 2014; Dargnies, 2012; Grosse & Riener, 2010; Healy & Pate, 2011; Ifcher & Zarghamee (IZ), forthcoming; Kamas & Preston, 2012; Niederle, Segal, & Vesterlund, 2013; Shurchkov, 2012; Sutter & Glätzle-Rützler, 2015; and Wozniak, Harbaugh & Mayr, 2014). A few studies, however, find that beliefs explain all of the gender gap (Cadsby, Servátka, & Song, 2013; Dreber, van Essen, & Ranehill, 2014; and Kamas & Preston, 2009). Researchers have also manipulated beliefs by presenting information about other subjects' performance (e.g., Cason, Masters, & Sheremeta, 2010; Dreber et al., 2014; Ertac & Szentes, 2010; and Wozniak et al., 2014), and by using real-effort tasks that are less stereotypically male (e.g., Grosse & Riener, 2010; Kamas & Preston, 2009; Shurchkov, 2012; and Wozniak et al., 2014).[2] These modifications reduce, and sometimes eliminate, the gender difference in the willingness to compete. However, it should be noted that NV chose the summation task purposefully, asserting that it should help us "understand why women are reluctant to enter competitive and male-dominated fields" (NV, 2011).

To examine the role of ***risk preferences***, some studies have followed the NV (2007) methodology and replicated the finding that its role is negligible (Dargnies, 2012; Healy & Pate, 2011; and Niederle et al., 2013). Other studies have controlled for risk preferences explicitly by eliciting them with incentivized lotteries and found, in line with NV, that risk preferences do not explain the residual gender gap (e.g., Almås et al., 2014; Buser et al., 2014; Cason et al., 2010; IZ, forthcoming; Kamas & Preston, 2012; Sutter & Glätzle-Rützler (2015), and Wozniak et al., 2014).

The residual-gender-gap explanation not considered in NV (2007) is ***other-regarding preferences***. This explanation is germane, as the WTA tournament is inherently less egalitarian than the

---

[2] In Niederle (forthcoming), the ***role of the task*** is considered separately from beliefs, while the two are considered together in NV (2011).

PR payment scheme.[3] Most studies that manipulate the payment distribution or account for social preferences find that it does not fully explain the residual gender gap. (e.g., Almås et al., 2014; Bartling et al., 2009; Cason et al., 2010; Dohmen & Falk, 2011; Kamas & Preston, 2009; and Teyssier, 2008). One study, however, finds that it does (Balafoutas et al., 2012). Taken together, these studies by and large confirm the NV (2007) conclusion that the gender gap in willingness to compete in a summation task is a product of variant beliefs and preferences for competition.[4]

In contrast to the extensive literature that tests the limits and robustness of NV (2007), we know of no study since that examines the relationship between the act of performance and the preference for competition. The assertions in NV (2007) imply an equivalence relationship between the preference for competition and taste for performing in a competitive environment; in other words, a taste for performing in a competitive environment is necessary and sufficient for a preference for competition. NV never use the language of necessity and sufficiency, though a claim of sufficiency is clearly indicated. For example, they conclude Section V.C.—entitled "Do Preferences for Performing in a Competition Cause Gender Differences in Choice of Compensation Scheme?"—by stating that their evidence "suggests that the gender gap in tournament entry is influenced by men and women differing in their preference for performing in a competitive environment (p. 1096)." Regarding necessity, NV show that, controlling for overconfidence, the gender gap with ex post choices is insignificant.[5] NV conclude, "When the compensation scheme choice does not require that participants subsequently perform in a competition, the relative overconfidence of men appears to explain most of the gender difference (p. 1094)." This can be understood to imply necessity.

Two concerns with NV's empirical evidence are as follows. First, they find a significant residual gender gap with ex ante but not ex post choices and interpret this as suggesting that the residual gender gaps are different. Rather, they have shown that there is a residual gender gap in one setting and not in the other. If the residual gender gap is not significantly larger with ex ante than ex post choices, then there is no evidence that performance is sufficient to generate a gender gap in the preference for competition.

---

[3] NV (2007) attempted to control for other-regarding preferences by designing the experiment so that subjects choose between a WTA or PR payment scheme for their own payments without imposing their chosen payment scheme on other subjects.

[4] Studies have also examined nature-versus-nurture bases for the gender gap in the willingness to compete. For example, it has been found that women are more willing than men to compete in a matrilineal society (Gneezy, Leonard, and List, 2009), and that the residual gender gap can be related to hormone-level variation from the menstrual cycle (e.g., Buser, 2012; and Wozniak et al., forthcoming). Sutter & Glätzle-Rützler (2015) find that the gender difference in willingness to compete is observable as early as kindergarten and persists over time. Other papers have found that institutional features— like affirmative action(e.g., Balafoutas & Sutter, 2012; and Niederle et al., 2013) or competing in groups rather than individually (e.g., Dargnies, 2012; and Healy and Pate, 2011)—can significantly reduce the residual gender gap.

[5] Of course, this is not the same as a necessity claim: the reported coefficient is negative, and it might not attain significance because of too small a sample size or the insensitivity of the willingness-to-compete measure.

Second, as discussed in NV, performance is not the only way in which ex ante and ex post choices differ. For example, with ex post choices, one knows the outcome before choosing to compete. This may impact one's risk preferences or feedback aversion so that they are not the same as in ex ante choices. Therefore, any difference in the residual gender gap with ex ante and ex post choices may not be attributable to performance.

In this paper, we provide additional evidence regarding both the sufficiency and necessity of performance for the gender gap in the preference for competition. We attempt to demonstrate sufficiency in two ways. First, we compare the residual gender gap with ex ante and ex post choices and find that the former is significantly larger than the latter. Second, we compare subjects' willingness to compete in the tasks with and without agency and find that the residual gender gap is significantly larger in the former. We test necessity with the residual gender gap in the task without agency and find that it is marginally significant, suggesting that performance may not be necessary for a gender gap in the preference for competition.

3. Experimental Design

Before presenting the experimental design, it should be noted that this experiment generated data for IZ (forthcoming) and this paper; for efficiency's sake, we designed the experiment to generate data that could be used to examine two disparate research questions. The focus of IZ (forthcoming) is validation of a new willingness-to-compete measure. IZ (forthcoming) introduces the new measure—designed to capture more information about willingness-to-compete than the binary-choice measure in NV (2007)— and examines willingness to compete in the summation task (described in section 3.2); Gneezy & Pietrasz (2013) and Petrie & Segal (2015) also develop measures designed to capture more information about willingness-to-compete than the binary-choice measure. All three studies find that the magnitude of men's willingness-to-compete is greater than women's and that the new measures are more informative than the binary-choice measure. The new willingness-to-compete measure validated in IZ (forthcoming) is utilized herein to examine a disparate research question: whether performing is necessary and sufficient to generate gender-variant preferences for competition. As such, IZ (forthcoming) does not consider the data generated from the task without agency (described in section 3.4) and restricts the sample to the 108 (of 208) subjects that completed the summation task first. Rather, this paper uses all the data.

Experimental sessions lasted approximately 90 minutes. Subjects received a minimum and average payment of $18 and $31, respectively. The experiment was programmed with the software z-Tree (Fischbacher, 2007). NV graciously provided their program, which we modified. All instructions were read aloud by the experimenter and appear in the Appendix.

3.1. Subjects

The experiment was conducted at Santa Clara University. 208 undergraduate students (109 male, 99 female) were recruited from required core courses in an attempt to generate a sample representative of the student body. Prospective subjects were told that the session would last approximately 90 minutes and that they would be paid for participating, with a minimum and average payment of $16 and $32, respectively. All subjects filled out informed consent forms at the start of the experiment. Subjects were given a pen and scrap paper but could not use calculators.

3.2. Tasks 1 – 5

This experiment builds upon NV (2007), and tasks 1 - 3 were designed to exactly replicate NV's tasks 1 - 3. In each of these tasks and in task 4, subjects had five minutes to sum sets of five randomly-chosen, two-digit numbers displayed horizontally across their screens (we denote tasks 1 – 4 as "tasks with agency" to emphasize that subjects had the ability to perform). After an answer was submitted, the computer indicated if the answer was correct, displayed the number of correct and incorrect answers submitted so far, and presented a new problem.[6] Subjects received no information regarding the performance of other subjects.

The payment scheme for tasks 1 – 4 varied. Before completing each task, subjects received detailed instructions regarding the task and payment scheme. The task-1 payment scheme was a $0.50 PR payment for each correct answer. The task-2 payment scheme was a WTA tournament. The subject with the most correct answers within their group of four received $2.00 per correct answer; other group members received no payment. Ties were settled randomly by the computer in all tasks. The instructions stated that a subject's group included the three subjects sitting in the same (front-to-back) row as the subject. The experimenter assigned subjects to seats by simply stating, "Please sit in seat 10." Seats were filled in a nonsystematic manner, except that the experimenter attempted to seat two female and two male subjects in each row. The groups' gender-balance was not discussed, however, subjects could observe it. Of 52 groups, 44 were balanced.[7]

In task 3, subjects were offered a choice between the $0.50 PR and $2.00 WTA payment. The instructions stated that if subjects choose the $2.00 WTA payment then their task-3 performance would be compared to their group-mates' task-2 performances. This ensured that subjects' task-3 payment-scheme choice could not impact other subjects' task-3 payments.

---

[6] During task 2 one subject reported that she could not submit her answer. The experimenter walked over to her computer and submitted the answer for her without incident. The subject did not report further problems. The experimenter told the subject that her payment would be adjusted if task 2 were selected for payment. Task 2 was not chosen for payment.

[7] The results are robust to including a dummy variable for unbalanced groups in the analysis.

In task 4, subjects were offered the payment scheme from IZ (forthcoming) that measures the willingness to compete along a continuum. Specifically, subjects were offered a series of choices between PR payments (ranging from 0 to 100 percent of the WTA payment) and the WTA payment. The first choice was between a $0.00 PR and $2.00 WTA payment. The second was between a $0.10 PR and $2.00 WTA payment. Thereafter, the PR payment increased in $0.10 increments until it reached $2.00. All choices were presented vertically on a single screen (see Figure 1 for a screenshot). We identify subjects' willingness to compete by observing their switch point: the minimum PR payment that they prefer to a $2.00 WTA payment. Switch points represent the PR-equivalent of the $2.00 WTA payment, hereafter denoted "PR-equivalent" (see IZ (forthcoming) for a discussion of PR-equivalents).[8]

Task 5 used the same payment scheme as task 4. However, subjects chose the payment scheme they wanted to apply to task 1 ex post. The number of correct answers in task 1 was displayed. No information regarding other subjects' performance in task 1 was displayed.

3.3. Beliefs regarding rank in tasks 1 – 4

For each of tasks 1 – 4, subjects were asked to rank themselves relative to their group-mates in terms of the number of correct answers they submitted. Subjects were informed that they would be paid $1 for each correct ranking.

3.4. Tasks 6 – 10

Tasks 6 - 10 were designed to replicate tasks 1 - 5, except the summation task was replaced with a task in which subjects had no agency. Specifically, subjects were informed that their computer would randomly generate numbers between one and four; that this was equivalent to their computer rolling a virtual, four-sided die; that their computer would roll the virtual die 20 times and display each roll for approximately ten seconds before displaying the next roll; that each roll was independent of all other rolls, so their rolls would not be the same as other subjects' rolls. Tasks 1 – 4 were entirely automated. The experimenter started the tasks, and subjects could not control or alter the trajectory of the tasks in anyway. There was no "start," "continue," or "end" button. Thus, subjects could take no actions. Rolling a 4 in the task without agency was equivalent to correctly answering a summation in the task with agency; subjects were not told this analogy. After each roll, the computer indicated if the roll was a 4 (or not) and displayed the count of rolls that were 4s (and not 4) so far. Subjects received no information regarding other subjects' rolls (see Figure 2 for a screenshot).

The payments for tasks 6 – 9 were analogous to tasks 1 – 4. Task 10 was analogous to task 5, with subjects making choices regarding the payment scheme for task 6 ex post. The number of 4s rolled

---

[8] As discussed in IZ (forthcoming), some subjects do not have a unique, identifiable switch point. For example, some switch between the WTA and PR payments more than once. Since no PR-equivalent can be inferred, the corresponding data are not used in corresponding analyses.

in task 6 was displayed, but no information regarding other subjects' rolls in task 6 was displayed. Lastly, 100 of the 208 subjects completed tasks 6 – 10 first, and tasks 1 – 5 second, enabling us to control for order effects in the regression analysis.[9]

3.5. Beliefs regarding rank in tasks 6 – 9

Subjects were asked to rank themselves relative to their group-mates in terms of the number of 4s rolled in tasks 6 – 9: 1st best, 2nd best, 3rd best, or 4th best. Subjects were informed they would be paid $1 for each correct ranking.

3.6. Task 11

Task 11 was a risk-preference measure in which subjects chose between a series of fixed payments, ranging from $0.00 to $10.00, and a lottery with even odds of getting either $10 or $0. All choices were presented vertically on a single screen. The first choice was between a $0.00 fixed payment and the lottery. The second was between a $1.00 fixed payment and the lottery. Thereafter, the fixed payment increased in $1.00 increments until it reached $10.00.

3.7. Exiting the Session

Subjects completed a questionnaire that included questions regarding demographic characteristics. Then subjects received their payments, which included a $5.00 show-up fee; $11.00 for completing all tasks; $1 for each correct ranking in tasks 1 – 4 and 6 – 9; and payment based on two tasks that were determined randomly at the end of the session using a bingo spinner. If tasks 4, 5, 9, or 10 were chosen, a PR payment was randomly chosen to be paid using the bingo spinner. If task 11 was chosen, a fixed payment was randomly chosen and the lottery was implemented using the bingo spinner. Subjects were paid in cash. The payments were placed in envelopes with only the subject's identification number on it. Subjects received their payment envelopes as they exited the session.

4. Results

Our task 3 replicates task 3 in NV (2007) and IZ (forthcoming). We find that men are more willing to compete than women in the summation task after controlling for performance, beliefs, and risk preferences. This residual gender difference is interpreted as indicating that men have a stronger preference for competition than women.[10] Also as in IZ (forthcoming), the residual gender gap is

---

[9] During the experiment, tasks 1 -5 and 6 - 10 were referred to as tasks 1s - 5s and 1r - 5r, respectively.
[10] Men and women average 10.8 and 10.6 correct answers (p = 0.72, Fisher's exact test: p = 0.84) (see Table 1). 57 and 37 percent of men and women (p = 0.01) choose the $2.00 WTA over $0.50 PR payment. We regress subjects' choice ($0.50 PR or $2.00 WTA payment) on gender, task-2 performance, task-1-to-2 performance improvement, task-2 self-rank, the lottery certainty equivalent, and a task-order dummy. The coefficient on female is negative and significant (probit marginal effect = -0.18, p = 0.02).

significant using the task-4 PR-equivalents.[11] We use PR-equivalents as the primary measure of willingness to compete, hereafter, as PR-equivalents provide more information regarding subjects' willingness to compete than the binary-choice measure in tasks 3 and 8.

To show that performing in a competitive environment generates an increase in the gender difference in the preference for competition, we first compare the residual gender gap with ex ante and ex post choices (this test was not conducted in NV (2007), where, instead, the two residual gender gaps were individually compared to zero). We regress subjects' task-4 (5) PR-equivalent on gender, task-3 (1) performance, task-2-to-3 performance improvement, task-3 (1) self-rank, the lottery certainty equivalent, and a task-order dummy. The task-4 and -5 coefficients on female are b = -0.28 (p = 0.00) and b = -0.10 (p = 0.11). A Wald test of the equality of the task-4 and -5 coefficients on female, using seemingly unrelated estimation clustering by subject, yields that they are significantly different (chi2(1) = 6.82, p = 0.01); this indicates that the gender difference in the preference for competition is greater with ex ante than ex post choices in the task with agency. As discussed above this is only suggestive of sufficiency, as subsequent performance is not the only difference between ex ante and ex post choices.

We provide additional evidence of sufficiency by comparing the willingness to compete in the tasks with and without agency. For ease of exposition, though, we first present the results of the task without agency, and in so doing, address the necessity claim (that there is a residual gender gap in the task without agency).

The distribution of task-9 PR-equivalents is illustrated in Figure 4. The mean PR-equivalent is $0.73 (SD = 0.40), indicating that on average subjects require a $0.73 PR payment to switch from the $2.00 WTA payment to a PR payment. The median PR-equivalent is $0.60. Examining the distribution of task-9 PR-equivalents by gender, one observes that the women's distribution is to the left of the men's (Fisher's exact test: p = 0.01). Men's mean PR-equivalent is significantly greater than women's ($0.80 and $0.65, p = 0.01). Men's median PR-equivalent is also greater than women's ($0.80 and $0.60). Finally, while only eight percent of men have a PR-equivalent less than $0.50, 23 percent of women do. In contrast, 21 percent of men and nine percent of women have a PR-equivalent greater than $1.00. In summary, men appear more willing to compete than women in the task without agency

To examine whether this gender difference is related to prior outcomes, beliefs, or risk preferences, we first explore if there are gender differences in these covariates, and then we estimate the residual gender difference in the willingness to compete. There is no gender difference in outcome and self-rank across tasks 6 – 9. The mean number of 4s is 5.0 for men and 4.9 for women (p = 0.58), and the

---

[11] Men and women average 10.9 and 10.6 correct answers (p = 0.62, Fisher's exact test: p = 0.32) (see Table 1). Women's distribution of PR-equivalents is to the left of men's (Fisher's exact test: p = 0.00) (see Figure 3). Men require a $1.00 PR payment to forgo the $2.00 WTA payment, while women require a $0.64 PR payment (p = 0.00).

mean self-rank (1st, 2nd, 3rd, or 4th best) is 2.4 for both men and women (p = 0.77) (see Table 1). Finally, men are less risk averse than women (mean lottery certainty equivalent: $5.20 and $4.70, p = 0.03).

That men and women reported the same self-rank in the task without agency diverges sharply with the task with agency. Men are more confident than women in the task with agency (mean self-rank: 1.9 and 2.3, p = 0.00), even though men and women perform equally well across tasks 1 – 4 (mean correct answers: 10.2 and 10.0, p = 0.72) (see Table 1). In other words, compared to the task with agency, the task without agency eliminated performance *de jure* and also eliminated the gender gap in overconfidence *de facto*. NV cite gender differences in overconfidence and preferences for performing in a competitive environment as the primary drivers of the gender gap in willingness to compete. Both have been eliminated in the task without agency. Given the minimal role that differences in risk preferences and feedback aversion have been shown to play in explaining the gender gap in willingness to compete, it is perhaps surprising that there is any gender difference in the task without agency.

To test necessity, we regress task-9 PR-equivalents on gender, task-8 outcome, task-7 to -8 improvement in outcome, task-8 self-rank, the lottery certainty equivalent, and a task-order dummy. The residual gender difference is negative and marginally significant (b = -0.10, p = 0.07) (see Column 1 of Table 2). This suggests that men may have a stronger preference for competition than women in the task without agency. This is true even when controlling for risk preferences, the only determinant in NV's theoretical model for which we observe a gender gap. Column 2 of Table 2 replicates the analysis of Column 1, excluding the lottery certainty equivalent; the residual gender gap is negative and significant (b = -0.16, p = 0.01) when measured risk preferences are not controlled for. Risk preferences, as measured by the certainty equivalent, account for approximately 38 percent (= 1 - (-0.10/-0.16)) of the gender gap reported in Column 2. The residual gender gap (reported in Column 1) is noteworthy, as task 9 was carefully designed to be performance-free and also happened to equalize overconfidence. Thus, performing—and gender-variant thrills or anxieties about performing in a competitive environment—may not be necessary to observe gender-variant preferences for competition. The residual gender gap may be driven by aspects of the competitive environment other than performing in it—for example, simply having a stake in a competitive process. Lastly, it appears that a gender gap in overconfidence is not necessary to observe gender-variant preferences for competition.[12]

---

[12] One could seek further evidence of necessity by testing the significance of the residual gender difference with ex post choices in either task, as there is no subsequent performance with ex post choices. In separate regressions of task-5 and -10 PR-equivalents on gender, outcome, outcome improvement, the lottery certainty equivalent, self-rank, and a task-order dummy, the residual gender differences are negative and insignificant (b = -0.10, p = 0.104, and b = -0.08, p = 0.14, respectively) (see Column 2 of Table 3). Recall that the task-9 residual gender difference is similar in magnitude (b = -0.10, p = 0.07). The task-5 and -9 residual gender differences are jointly marginally significant (p = 0.09) and the task-5, -9, and 10 residual gender differences are similar in magnitude and statistically indistinguishable. Together, this evidence might suggest that these residual gender differences—all in the absence

Returning to the claim of sufficiency, to examine whether performance generates a gender difference in both the willingness to compete and preference for competition, we compare subjects' PR-equivalents in the tasks with and without agency. First, we separately regress task-4 and -9 PR-equivalents on gender and a task-order dummy and perform a Wald test of the equality of the task-4 and -9 coefficients on female using a seemingly unrelated estimation clustering by subject (see Table 4). Second, we repeat the analysis with the residual gender gap (adding controls for outcome, improvement in outcome, self-rank, and the lottery certainty equivalent) (see Table 5). In Table 4, the task-4 and -9 coefficients on female are negative and significant (b = -0.37, p = 0.00, and b = -0.15, p = 0.01), indicating that men are more willing to compete than women in both tasks. Further, and speaking to sufficiency, the coefficients are not equal (chi2(1) = 7.23, p = 0.01), suggesting that performance generates gender-variant willingness to compete. In Table 5, the task-4 and -9 coefficients on female are negative and at least marginally significant (b = -0.28, p = 0.00, and b = -0.10, p = 0.07, respectively), suggesting that men have a stronger preference for competition than women in both tasks. Again speaking to sufficiency, the coefficients are not equal (chi2(1) = 5.00, p = 0.03), suggesting that performance generates gender-variant preference for competition.

5. Discussion

We provide evidence that ***performing*** in a competitive environment is sufficient, but might not be necessary, to observe a gender difference in the preference for competition. Our findings vis-a-vis sufficiency suggest that performance, and the thrills and anxieties it generates, impact the preference for competition, and that the well-established gender-gap in willingness to compete cannot be eliminated by correcting individuals' beliefs, alone. Our findings vis-a-vis necessity suggest that the elimination of performance—for example, by routinizing tasks—may not eliminate the gender gap in the preference for competition, and that this gender gap must have other determinants. Below, we will discuss sufficiency and necessity, in turn.

We utilize a more sensitive willingness-to-compete measure, as developed in IZ (forthcoming), to establish that the gender gap in the preference for competition is stronger with ex ante than ex post choices. This suggests that *subsequent* performance generates a gender gap in preference for competition, but this evidence is inadequate to establish sufficiency because of other possible differences between ex ante and ex post choices.

---

of subsequent performance—fall short of significance because of a lack of power and not because they are equal to zero.

Therefore we create a novel task without agency and find that gender differences in willingness to compete and the preference for competition are greater in the task with than without agency. The task without agency is designed to retain the numerical nature of NV's summation task, but to eliminate performance (decision-making and actions). Further, the task without agency was designed to minimize any *illusion of control*, whereby subjects might think that their actions influence random outcomes. For example, subjects cannot take even meaningless actions, such as, clicking an on-screen "button" to reveal a pre-determined number. That self-rank in the task without agency was well-calibrated for both men and women—both guessed self-rank of 2.4 on average (the expected rank is 2.5)—suggests that illusion of control was indeed minimized.

The two residual-gender-gap comparisons (between ex ante and ex post choices and between tasks with and without agency) that we use to establish sufficiency each have their pros and cons. In the comparison of ex ante and ex post choices, the task is held constant, but there is variation in information about the task outcome and feelings about the competitive *environment* (feelings unrelated to subsequent performance—for example, preference for the chance to win over others or aversion to the chance of losing to others). In contrast, in the comparison of tasks with and without agency, the task varies, but information about the task outcome and feelings about the competitive environment are similar.

We have suggestive evidence that the competitive environment varies with ex ante and ex post choices. Comparing the coefficients on the controls from the regressions with ex ante and ex post choices for both tasks, one finds significant differences (see Table 6). For example, in the task with agency, the coefficients on outcome improvement and self-rank are both at least marginally different. In contrast, comparing the coefficients on the controls from the regressions with ex ante choices for both tasks, there are no significant differences. We also have suggestive evidence that varying the tasks varies more than just the ability to perform. For example, the de facto elimination of gender differences in self-rank in the task without agency means that performance and gender-variant overconfidence are *together* sufficient for gender-variant preferences for competition. That said, we control for self-rank in the comparison of the tasks with and without agency, so only unobserved overconfidence would confound our result.

While both the ex ante/ex post and with/without agency comparisons have their weaknesses, taken together, they provide solid evidence of sufficiency. Any alternate explanation arising from a weakness of one comparison is unlikely to explain the result vis-a-vis the other. For example, the residual gender difference may be greater in the task with than without agency because men have a stronger preference than women for "deserved" than "undeserved" winning. However, "deservedness" would presumably be similar with ex ante and ex post choices in the task with agency.

We now turn to a discussion of the necessity claim. In the task without agency, we find a significant unconditional gender difference and a marginally significant (p = 0.07) residual gender

difference (controlling for outcome, beliefs, and risk preferences). Recall, men are not more confident than women in the task without agency, so the result cannot be driven by gender differences in overconfidence. The unconditional result is perhaps unsurprising, as previous research has shown that men are more risk-loving than women. The residual gender difference, however, explicitly controls for measured risk preferences, yet there is still evidence of a gender difference. Two other potential explanations do not appear plausible either: feedback aversion and other-regarding preferences. Each are present with ex post choices in both tasks; and the residual gender differences are similar in magnitude and statistically indistinguishable.

Given the above and, again, that the task without agency eliminates performance de jure and overconfidence de facto, what might be driving gender differences in the preference for competition in the task without agency? Three plausible explanations emerge. First, risk preferences may not be fully measured by the instrument utilitzed. Consequently, a gender difference in unmeasured risk preferences could be responsible for the residual gender gap in the preference for competition with ex ante choices in the task without agency.[13] Second, there may be heretofore unidentified determinants of the preference for competition. For example, as mentioned above, men may prefer being subject to payment schemes that are based on social comparison or in which there is a chance to win over others. Lastly, men may prefer such payment schemes per se, even when the payment scheme is applied to others and not themselves.

---

[13] It is also important to consider how willingness to compete in the task without agency differs from risk preferences. First, in our experiment, subjects are informed that they are competing head-to-head against three group-mates in the task without agency and that they will receive the WTA payment only if their outcome is greater than their group-mates' outcomes. Second, the subject is the object of his or her own risk-taking as opposed to the object being a general, non-personalized random event. In other words, each subject is both the subject *and* object of competition, which is different from, for example, betting on a sports game or lottery.

5. References

Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen, and Bertil Tungodden. (2014). "Willingness to compete: Family matters." Department of Economics, NHH Discussion Papers SAM 3/2014.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter. (2012). "Distributional preferences and competitive behavior." *Journal of Economic Behavior & Organization,* 83(1): 125-135.

Balafoutas, Loukas, and Matthias Sutter (2012). "Affirmative action policies promote women and do not harm efficiency in the laboratory." *Science,* 335: 579-582.

Bartling, Björn, Ernst Fehr, and Daniel Schunk. (2012). "Health effects on children's willingness to compete." *Experimental Economics*: 58-70.

Buser, Thomas. (2012). "The impact of the menstrual cycle and hormonal contraceptives on competitiveness." *Journal of Economic Behavior & Organization,* 83(1): 1-10.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. (2014). "Gender, competitiveness and career choices." *Quarterly Journal of Economics,* 129(3): 1409-1447.

Cadsby, C. Bram, Maroš Servátka, and Fei Song. (2013). "How competitive are female professionals? A tale of identity conflict." *Journal of Economic Behavior & Organization,* 92: 284-303.

Cason, Timothy N., William A. Masters, and Roman M. Sheremeta. (2010). "Entry into winner-take-all and proportional-prize contests: An experimental study." *Journal of Public Economics,* 94(9): 604-611.

Dargnies, Marie-Pierre. (2012). "Men too sometimes shy away from competition: The case of team competition." *Management Science*, 58(11): 1982-2000.

Dohmen, Thomas, and Armin Falk. (2011). "Performance pay and multidimensional sorting: Productivity, preferences, and gender." *American Economic Review*, 101(2): 556-590.

Dreber, Anna, Emma von Essen, and Eva Ranehill (2014). "Gender and competition in adolescence: task matters." *Experimental Economics*, 17(1): 154-172.

Ertac, Seda, and Balazs, Szentes. (2010). "The effect of performance feedback on gender differences in competitiveness: experimental evidence." Working Paper, Koc University, Turkey.

Fischbacher, Urs. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2): 171-178.

Gneezy, Uri, Kenneth L. Leonard, and John A. List. (2009). "Gender differences in competition: Evidence from a matrilineal and a patriarchal society." *Econometrica*, 77(5): 1637-1664.

Gneezy, Uri, and Aniela Pietrasz. (2013). "When half of the men are more competitive than all women." Working paper.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. (2006). "The homecoming of American college women: The reversal of the college gender gap." *Journal of Economic Perspectives*, 20(4): 133-156.

Große, Niels Daniel, and Gerhard Riener. (2010). "Explaining gender differences in competitiveness: gender-task stereotypes." No. 2010, 017. Jena Economic Research Papers.

Healy, Andrew, and Jennifer Pate. (2011). "Can Teams Help to Close the Gender Competition Gap?" *Economic Journal*, 121(555): 1192-1204.

Ifcher, John, and Homa Zarghamee. (forthcoming). "Pricing competition: A new laboratory measure of gender differences in the willingness to compete." *Experimental Economics*.

Kamas, Linda, and Anne Preston. (2009). "Social preferences, competitiveness and compensation: Are there gender differences?" Santa Clara University Working Paper.

Kamas, Linda, and Anne Preston. (2012). "The importance of being confident; gender, career choice, and willingness to compete." *Journal of Economic Behavior & Organization*, 83(1): 82-97.

Niederle, Muriel. (forthcoming). "Gender." In preparation for the *Handbook of Experimental Economics, Vol.2,* Editors: Kagel, John, and Alvin E. Roth.

Niederle, Muriel, Carmit Segal, and Lise Vesterlund. (2013). "How costly is diversity? Affirmative action in light of gender differences in competitiveness." *Management Science*, 59(1): 1-16.

Niederle, Muriel, and Lise Vesterlund. (2007). "Do women shy away from competition? Do men compete too much?" *Quarterly Journal of Economics*, 122(3): 1067-1101.

Niederle, Muriel, and Lise Vesterlund. (2011). "Gender and competition." *Annual Review of Economics*, 3(1): 601-630.

Petrie, Ragan, and Carmit Segal. (2015). "Gender differences in competitiveness: The role of prizes." George Mason University Working Paper in Economics No. 14-47.

Sutter, Matthias, and Daniela Glätzle-Rützler (2015). "Gender differences in the willingness to compete emerge early in life and persist." *Management Science*, 61(10): 2339-2354.

Teyssier, Sabrina. (2008). "Experimental evidence on inequity aversion and self-selection between incentive contracts." GATE Work. Pap. 08-21.

Wozniak, David, William Harbaugh, and Ulrich Mayr. (2014). "The menstrual cycle and performance feedback alter gender differences in competitive choices." *Journal of Labor Economics*, 32(1): 161-198.

Figure 1: Screenshot of willingness-to-compete measure

Figure 2: Screenshot of task without agency

Roll number	4
The computer rolled a	1

This roll	was not a 4
Rolls that WERE NOT a 4 so far	2
Rolls that WERE a 4 so far	2

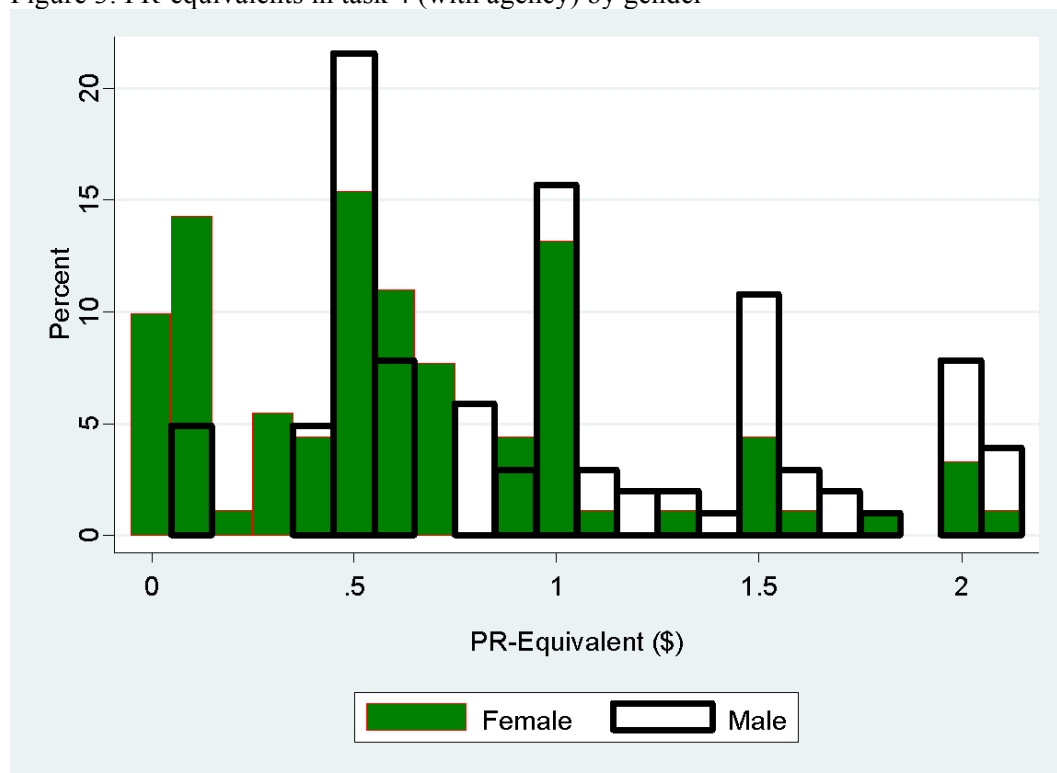Figure 3: PR-equivalents in task 4 (with agency) by gender

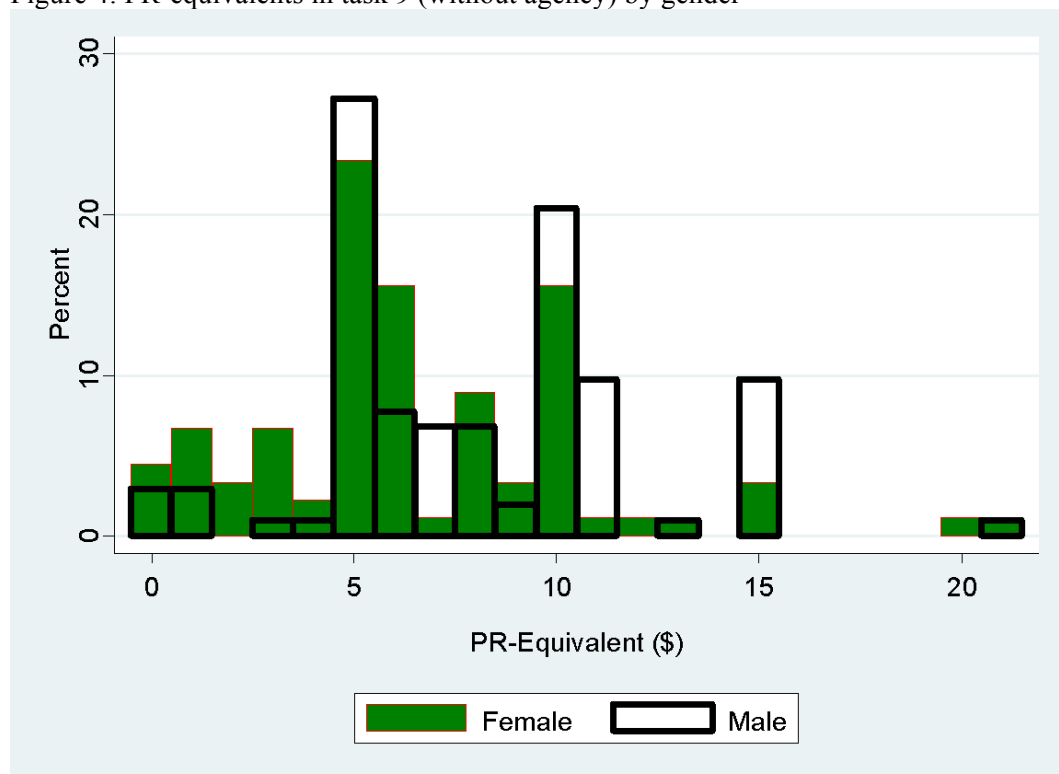Figure 4: PR-equivalents in task 9 (without agency) by gender

Table 1: Outcomes and self-ranks in tasks with and without agency

	Task with agency (Tasks 1 - 4)			Task without agency (Tasks 6 - 9)		
	Men (1)	Women (2)		Men (3)	Women (4)	
Task-1 (6) outcome	8.936 (0.465)	8.646 (0.280)		4.936 (0.201)	4.687 (0.184)	
Task-2 (7) outcome	10.183 (0.478)	10.202 (0.311)		4.771 (0.194)	5.131 (0.224)	
Task-3 (8) outcome	10.835 (0.497)	10.616 (0.313)		5.000 (0.188)	5.152 (0.188)	
Task-4 (9) outcome	10.945 (0.504)	10.636 (0.333)		5.110 (0.179)	4.535 (0.194)	**
Mean outcome	10.225 (0.461)	10.025 (0.268)		4.954 (0.096)	4.876 (0.105)	
Task-1 (6) self-rank	2.202 (0.089)	2.677 (0.086)	***	2.376 (0.106)	2.505 (0.097)	
Task-2 (7) self-rank	1.844 (0.089)	2.182 (0.091)	***	2.477 (0.102)	2.364 (0.101)	
Task-3 (8) self-rank	1.716 (0.081)	2.162 (0.092)	***	2.257 (0.107)	2.222 (0.097)	
Task-4 (9) self-rank	1.881 (0.093)	2.192 (0.104)	**	2.339 (0.106)	2.424 (0.108)	
Mean self-rank	1.911 (0.060)	2.303 (0.066)	***	2.362 (0.038)	2.379 (0.040)	

NOTES: Standard errors are in parentheses. *, **, and *** signify that the means in are significantly different for men and women with a p-value < 0.10, 0.05, and 0.01, respectively.

Table 2: Estimates from regressing task-9 PR-equivalents on gender, outcome, self-rank, and certainty equivalent

	Task-9 (without agency) PR-equivalent (1)			Task-9 (without agency) PR-equivalent (1)		
Female	-0.097	(0.054)	*	-0.157	(0.057)	***
Prior-task outcome	0.011	(0.019)		0.021	(0.022)	
Improvement in outcome (between 2 prior tasks)	-0.018	(0.012)		-0.020	(0.015)	
Prior-task self-rank (1st, 2nd, 3rd, & 4th best)	-0.030	(0.029)		-0.034	(0.031)	
Lottery certainty equivalent (risk preference)	0.110	(0.016)	***			
Task-order dummy	0.005	(0.053)		0.024	(0.058)	
Observations	191			191		

NOTES: Standard errors are in parentheses. *, **, and *** signify that ceofficient is significantly different from zero with a p-value < 0.10, 0.05, and 0.01, respectively.

Table 3: Comparing the preference for competition with ex ante and ex post choices

Gender difference in preference for competition	Ex ante choice (1)			Ex post choice (2)		
Task with agency (A)	-0.275	(0.068)	***	-0.100	(0.061)	^^^
Task without agency (B)	-0.097	(0.054)	*	-0.083	(0.056)	
	^^					

NOTES: Standard errors are in parentheses. *, **, and *** signify that coefficient is significantly different from zero with a p-value < 0.10, 0.05, and 0.01, respectively. ^, ^^, ^^^ signify that the Columns (1) and (2) - or rows (A) and (B) - are significantly different from each other with a p-value < 0.10, 0.05, and 0.01, respectively.

Table 4: Comparing the willingness to compete in the tasks with and without agency

	Task-4 (with agency) PR-equivalent (1)			Task-9 (without agency) PR-equivalent (2)			
Female	-0.368	(0.078)	***	-0.149	(0.057)	***	^^^
Task-order dummy	0.053	(0.078)		0.040	(0.057)		
Observations	193			193			

NOTES: Standard errors are in parentheses. *, **, and *** signify that ceofficient is significantly different from zero with a p-value < 0.10, 0.05, and 0.01, respectively. ^, ^^, ^^^ signify that the Columns (1) and (2) are significantly different from each other with a p-value < 0.10, 0.05, and 0.01, respectively.

Table 5: Comparing the preference for competition in the tasks without and with agency

	Task-4 (with agency) PR-equivalent (1)			Task-9 (without agency) PR-equivalent (2)			
Female	-0.275	(0.068)	***	-0.097	(0.054)	*	^^
Prior-task outcome	0.034	(0.008)	***	0.011	(0.019)		
Improvement in outcome (between 2 prior tasks)	-0.044	(0.017)	***	-0.018	(0.012)		
Prior-task self-rank (1st, 2nd, 3rd, & 4th best)	-0.092	(0.047)	*	-0.030	(0.029)		
Lottery certainty equivalent (risk preference)	0.101	(0.025)	***	0.110	(0.016)	***	
Task-order dummy	0.047	(0.068)		0.005	(0.053)		
Observations	192			191			

NOTES: Standard errors are in parentheses. *, **, and *** signify that ceofficient is significantly different from zero with a p-value < 0.10, 0.05, and 0.01, respectively. ^, ^^, ^^^ signify that the Columns (1) and (2) are significantly different from each other with a p-value < 0.10, 0.05, and 0.01, respectively.

Table 6: Comparing the preference for competition with ex ante and ex post choices

	Task-4 (with agency and ex ante) PR-equivalent (1)	Task-5 (with agency and ex post) PR-equivalent (2)	Task-9 (without agency and ex ante) PR-equivalent (3)	Task-10 (without agency and ex post) PR-equivalent (4)
Female	-0.275 (0.068) ***	-0.100 (0.061) ^^^	-0.097 (0.054) *	-0.083 (0.056)
Prior-task outcome	0.034 (0.008) ***	0.032 (0.008) ***	0.011 (0.019)	0.052 (0.025) **
Improvement in outcome	-0.044 (0.017) ***	-0.003 (0.012) ^^	-0.018 (0.012)	0.021 (0.012) * ^^
Self-rank	-0.092 (0.047) *	-0.199 (0.038) *** ^	-0.030 (0.029)	-0.266 (0.042) *** ^^^
Lottery certainty equivalent	0.101 (0.025) ***	0.074 (0.022) ***	0.110 (0.016) ***	0.046 (0.021) ** ^^
Task-order dummy	0.047 (0.068)	0.112 (0.060) *	0.005 (0.053)	0.037 (0.060)
Observations	192	192	191	191

NOTES: Standard errors are in parentheses. *, **, and *** signify that ceofficient is significantly different from zero with a p-value < 0.10, 0.05, and 0.01, respectively. ^, ^^, ^^^ signify that the Columns (1) and (2) - or (3) and (4) - are significantly different from each other with a p-value < 0.10, 0.05, and 0.01, respectively.