NEW YORK ECONOMIC REVIEW

JOURNAL OF THE NEW YORK STATE ECONOMICS ASSOCIATION

VOLUME LIII

Vol. 53, Fall 2023

CONTENTS

EDITORIAL

The *New York Economic Review* (NYER) is an annual publication of the New York State Economics Association (NYSEA). The NYER publishes theoretical and empirical articles, and also interpretive reviews of the literature in the fields of economics and finance. All well-written, original manuscripts are welcome for consideration at the NYER. We also encourage the submission of short articles and replication studies. Special Issue proposals are welcome and require a minimum of 4 papers to be included in the proposal, as well as a list of suggested referees.

MANUSCRIPT GUIDELINES

All manuscripts are to be submitted via https://www.nyseconomicsassociation.org/new-york-economic-review/submit-paper-to-the-editor/ to the editor in a Microsoft Word file format, with content formatted according to the guidelines at:

https://www.nyseconomicsassociation.org/NYER/Guidelines

The NYER is cited in:

- EconLit by the American Economic Association
- Ulrich's International Periodicals Directory
- Cabell's Directory of Publishing Opportunities in Business and Economics
- Australian Business Deans Council (ABDC) Journal Quality List

ISSN NO: 1090-5693

Completed Extensions and Set-Theoretical Structures of an Individual's Consumption Preferences

Jeffrey Yi-Lin Forrest[*], Lawrence Shao[†], Shynara Sarkambayeva (Jumadilova)[‡], Dale Shaw[§], Sunita Mondal[**]

ABSTRACT

There is evidence that consumers have incomplete and intransitive consumption preferences. This paper examines if a preference relation that is only reflexive can be extended to a sensually complete binary relation. Our affirmative conclusion is similar to Szpilrajn extension theorem for partial orders and Hansson theorem for preorders. As consequences, we show that the order dimension of each incomplete partial order or incomplete preference relation is equal to 2; and that each countable, incomplete and partially ordered set $(X, \precsim)$ is representable by a function $u: X \to \mathbb{R}^2$. All of these results majorly improve a series of conclusions derived in the recent past.

1. INTRODUCTION

Various empirical evidences and theoretical reasonings clearly point to the fact that for any consumer, his preferences of commodity consumption generally only satisfy the property of reflexivity (e.g., Aumann, 1962; Birnbaum & Gutierrez, 2007; Bosi & Herden, 2012; Cettolin & Riedl, 2019; Dubra & Ok, 2002; Evren & Ok, 2011; Hansson, 1968; Mandler, 1999; Nishimura & Ok, 2016; Ok, 2002; Tversky, 1969). However, for the established consumer theory (Levin & Milgrom, 2004a; Mas-Colell et al., 1995; Miller, 2006) and choice theory (e.g., Glasser, 1999; Levin & Milgrom, 2004b), many important conclusions are developed on such preference relations that are assumed to also satisfy transitivity and completeness. So, it is both theoretically and practically important to see how these conclusions would look like when the preference relation of concern is merely reflexive without transitivity and completeness. To help possibly accomplish this end, this paper investigates how some of the fundamental results regarding partial orders or preorders can be established for preference relations that are reflexive without definitely satisfying the conditions of transitivity and completeness.

Specifically, this paper looks at the question of whether or not a result similar to Szpilrajn extension theorem for partial orders and Hansson theorem for preorders can also hold true for preference relations without transitivity and completeness. Based on our affirmative answer to this question, we are able to

[*]Department of Accounting Economics Finance, Slippery Rock University, PA, Email: Jeffrey.forrest@sru.edu
[†]College of Business, Slippery Rock University, PA
[‡]Department of Management and Mathematical Economics, Satbayev University, Kazakhstan
[§]Lewis College of Business, Marshall University, WV
[**]Department of Accounting Economics Finance, Slippery Rock University, PA

show, amongst others, that (1) each such preference relation is equal to the intersection of all completed consumption preferences; (2) each such relation $\gtrsim$ has two completed extensions L_1 and L_2 so that for each pair of $\gtrsim$-incomparable $x, y \in X$, either (xL_1y and yL_2x) or (yL_1x and xL_2y), but not both, holds true; and (3) the order dimension of every such relation (and that of each incomplete partial order) is equal to 2. And each of the developed results in this paper greatly improves corresponding results established in the past by different authors.

The rest of the paper is organized as follows. Section 2 familiars the reader with the basic terminology used in the rest of this presentation. Section 3 derives the result that each preference relation can be completely extended so that originally incomparable consumptions can now be compared. Section 4 looks at the order dimension of preference relations. Section 5 pays a revisit to linear extensions of partial orders. Then, this paper is concluded in Section 6.

2. SOME RELATED TERMINOLOGY OF A CONSUMER'S CONSUMPTION PREFERENCES

Let X be the set of all possible consumptions of the focal consumer. As in Debreu (1959), let X contain all possible consumptions of the consumer in his entire lifetime without discounting the future. So, X will be seen as an infinite set. For $x, y \in X$, if the consumer can choose only one of them, then in real life he would likely choose the one that is more preferred than the other. In such a case, consumptions x, y are said to be comparable with each other in terms of the consumer's preferences and satisfy one and only one of the following alternatives: (1) x is preferred to y; (2) x is indifferent to y; (3) y is preferred to x.

Let $\succ$ denote the focal consumer's preference relation (or simply preference) of the consumptions in X. Then, $\succ$ is a binary relation defined on X. That is, $\succ$ is a subset of $X \times X$. If $(x, y) \in X \times X$ belongs to $\succ$, we use $x \succ y$ to mean that x is preferred to y, and $x \gtrsim y$ to mean that x is at least as preferred as to y. Two possible consumptions $x, y \in X$ are $\gtrsim$-comparable, if either x $\gtrsim$ y or y $\gtrsim$ x holds; otherwise x and y are said to be $\gtrsim$-incomparable. Such a binary relation $\gtrsim$ is said to be complete if any possible consumptions x and y in X are $\gtrsim$-comparable, otherwise incomplete. The strict or asymmetric part of the binary relation $\gtrsim$ is denoted as $\succ$, representing a relation on X defined as $x \succ y$ iff $x \gtrsim y$ and $\neg(y \gtrsim x)$, for $x, y \in X$. The symmetric part of the relation $\gtrsim$ represents the relationship of indifferences, denoted as $\sim$, and is defined as follows: for $x, y \in X$, $x \sim y$ iff $x \gtrsim y$ and $y \gtrsim x$. Symbolically, $\sim \equiv \gtrsim \setminus \succ$. In other words, $x \sim y$ means that consumptions x and y are indifferent in terms of the consumer's consumption preferences.

This preference relation $\gtrsim$ on X is said to be a preorder (or a quasiorder), provided that $\gtrsim$ is reflexive and transitive; that is, for any consumptions $x, y, z \in X$, $x \gtrsim x$ and $(x \gtrsim y) \wedge (y \gtrsim z) \rightarrow (x \gtrsim z)$. It is a partial order, if, in addition to reflexivity and transitivity, $\gtrsim$ satisfies antisymmetry; that is, for any $x, y \in X$, $x \gtrsim y$ and $y \gtrsim x$ imply $x = y$. It is a linear order, provided that $\gtrsim$ is a complete partial order; that is, $\gtrsim$ is a partial order and for any $x, y \in X$, x, y are $\gtrsim$-comparable.

With these terms in place, it can be readily seen that the focal consumer's preference of consumptions can be at most a preorder, as the situation in real life. First, because of the existence of indifferent consumptions, antisymmetry required for a partial order cannot hold for consumption preferences in real life. As for transitivity, Aumann (1962) and Mandler (1999) find that due to wide range appearances of indecisiveness, consumer preferences tend to be intransitive, while Tversky (1969) reports that consumer preferences don't generally satisfy the condition of transitivity. And, by using a new statistical technique and by revisiting the same gambles Tversky studied earlier, Birnbaum and Gutierrez (2007) conclude that there are indeed individual consumers who repeat intransitive preference patterns. So, in the rest of this paper, consumption preferences are not assumed to be preorders; and without being particularly specified, each consumption preference is only assumed to be reflective. That is, when a binary relation R on a set X is said to be a consumption preference (or simply preference), it means that R is reflexive. For related, but dissimilar works, see, for example, Nishimura and Ok (2016) and references found there.

Because each individual consumer is a physiological being and each business consumer stands for a form of life (Lin & Forrest, 2011), the person's or firm's needs for basic survival and for better living conditions have to be multi-dimensional. That is, possible consumptions of any consumer, be it an individual or a firm, cannot be assumed to be completely comparable in terms of his preferences, as so commonly done in studies of economics (e.g., Debreu, 1959; Levin & Milgrom, 2004a; Mas-Colell et al., 1995). Speaking differently, when faced with two commodities from two different dimensions of survival, such as the dimension of shelter, that of foods, that of drinks, that of medicine, etc., no matter who is concerned with, he, as a consumer, cannot really say which commodity is preferred to the other. In terms of literature, several scholars had also noticed this issue of incompleteness in consumption preferences. For example, Dubra and Ok (2002) introduce a risky-choice model in which an individual naturally possesses an incomplete preference relation. Ok (2002), Nishimura and Ok (2016), and Bosi and Herden (2012) consider the problem of how to represent an incomplete preference relation by means of a collection of real-number valued functions. Based on such a quickly expending literature, Alonso et al. (2010) present a web-based consensus support system that involves decisions makers with incomplete preference relations; Meng and Chen (2015) develop a group-decision-making method to cope with incomplete preference information; and Cettolin and Riedl (2019) conduct experiments to test whether a preference is either complete or incomplete.

In the rest of this paper, without any particular emphasis, each preference relation is assumed to be merely reflexive without necessarily satisfying transitivity and completeness.

3. CONSUMPTION PREFERENCE AND ITS COMPLETED EXTENSIONS

Let R and S be two binary relations on a set X, satisfying reflexivity. The former relation R is said to be contained in the latter relation S, if $R \subset S$. That is, for any $x, y \in X$, xRy implies xSy, while there are $a, b \in X$ such that aSb and $\neg(aRb)$. In this case, S is seen as an extension of R. Corresponding to Szpilrajn extension theorem for partial orders (1930) and Hansson extension theorem for preorders (1968), one would naturally inquire the possibility to extend any given preference relation $\succsim$. Indeed, in pure symbolic terms, the answer to this inquiry is YES, because $\succsim$ can be trivially extended to $\succsim^*$ by treating all $\succsim$-incomparable pairs of possible consumptions as indifferent in $\succsim^*$. That is, $\succsim^* = \succsim \cup \{(x, y) \in X^2 \colon \neg x \succsim y \text{ and } \neg y \succsim x\}$.

The reason why we say that $\succsim^*$ exists in pure symbolic terms is because $\succsim$ is assumed to be a preference relation on the set X of all possible consumptions of our focal consumer. That is, other than being a reflexive binary relation on an abstract set $\succsim$ also embodies other meanings of life, such as the multidimensionality of a life's physiological needs. In particular, when x, y are two possible consumptions in X and incomparable in terms of the preference $\succsim$, it means that the focal consumer can in some way consume either x or y, while the incomparability between x and y indirectly implies that these consumptions potentially come from different physiological dimensions. Therefore, to make our established results practically useful instead of some additional symbolic statements irrelevant to rea life, such possible consumptions x and y cannot be and should not be simply treated as indifferent. Instead, they are different, very different from each other. They are indispensable for the survival and livelihood of the focal consumer. Due to this reason, any extension of the preference relation $\succsim$ that makes a pair of originally incomparable consumptions indifferent, such as the previous extension $\succsim^*$, will be referred to as non-sensual and denoted as $\succsim^{nonsense}$. Therefore, the natural inquiry about the possibility to extend a preference relation becomes that of finding a sensual extension, denoted as $\succsim^{sense}$, in terms of the physiological needs of a life form.

Theorem 1. Assume the Axiom of Choice and that each infinity can be actually (not potentially) achieved. Let $\succsim$ be the preference relation on the set X of all possible consumptions of the focal consumer. Then $\succsim$ can be extended to a sensual binary relation $\succsim^{sense}$ such that for any $x, y \in X$, x and y are $\succsim^{sense}$-comparable.

Before anything else, let us first clarify the meaning of the assumption. The concept of infinity involves both actual and potential infinities (Lin, 2008). A potential infinity means a present, progressive tense or a forever, ongoing and never-ending process; and every actual infinity represents a present or past perfect tense or a process that actually ends or had ended. It is found that these two different kinds of infinities do potentially lead to completely opposite answers (Lin, 2002), although in modern mathematics these infinities are treated either the same or different depending on the needs in hand (Forrest, 2013)[1].

Although the proof of this result, which is given in the appendix, is very technical, the idea underneath the argument is quite straightforward. First, by using the axiom of choice, well order X into $\{x^\alpha : \alpha \in I\}$, where I is the set of ordinal numbers in $|X|$ – the cardinality of X, as an index set[2], so that any nonempty subset of X has a unique element with the least index α. Define a subset N^0 of X by

$$N^0 = \{x \in X : \exists y \in X\, [\neg(x \succsim y) \wedge \neg(y \succsim x)]\}. \tag{2}$$

If x^{β_0} is the element in N^0 with the least index β_0, then we extend the preference relation $\succsim$ to

$$\succsim^0 = \succsim \cup \left\{\left(x^{\beta_0}, y\right) : y \in X\left[\neg\left(x^{\beta_0} \succsim y\right) \wedge \neg\left(y \succsim x^{\beta_0}\right)\right]\right\},$$

where $(x^{\beta_0}, y) \in \succsim^0$ mean that x^{β_0} is preferred over y in terms of the extended preference $\succsim^0$. Next, repeat this process by looking at

$$N^1 = N^0 - \{x^{\beta_0}\}$$

and extend $\succsim^0$ to $\succsim^1 = \succsim^0 \cup \{(x^{\beta_1}, y): y \in X\ [\neg(x^{\beta_1} \succsim y) \wedge \neg(y \succsim x^{\beta_1})]\}$, where x^{β_1} is the element in N^1 with the least index β_1 and $(x^{\beta_1}, y) \in \succsim^1$ mean that x^{β_1} is preferred over y in terms of the extended preference $\succsim^1$. When this process is completed, which is guaranteed by the transfinite induction applied on $\alpha \in I$ and the assumption that each infinity can be actually (not potentially) achieved, by checking through all the elements in $X = \{x^\alpha: \alpha \in I\}$, the desired sensual extension of $\succsim$ is then given by $\succsim^* = \succsim \cup \succsim^0 \cup \succsim^1 \cup \ldots$ For the detailed proof of this theorem, please go to the Appendix.

Due to the reason that a preference relation considered in this research does not require transitivity, the so-called Suzumura-consistency (Suzumura, 1976) is not needed for Theorem 1 to hold true. Here, a binary relation $\succsim$ is Suzumura-consistent if there are no such consumption choices x_i, $i = 0, 1, 2, \ldots, n-1$, for some natural number n, such that $x_i \succsim x_{i+1}$, for $i = 0, 1, \ldots, mod(n)$, and $\neg x_{j+1} \succsim x_j$, for some $j = 0, 1, \ldots, mod(n)$. And without causing confusion, in the rest of this paper, it is always assumed that that each infinity can be actually (not potentially) achieved.

4. ORDER DIMENSIONS OF PREFERENCE RELATIONS

Related to Theorem 1, Dushnik and Miller (1941) prove that any strict partial order is the intersection of strict linear orders; and Donaldson and Weymark (1998) confirm that each preorder is equal to the intersection of complete preorders. Accordingly, in terms of consumption preferences, we have the following result, where each binary relation $\succsim^+$, which sensually extends the consumption preference $\succsim$ so that for any $x, y \in X$, x and y are $\succsim^+$-comparable, is referred to as a sensually completed consumption preference or sensually completed extension.

Theorem 2. The preference relation $\succsim$ defined on the consumption set X of the focal consumer is equal to the intersection of all sensually completed consumption preferences of $\succsim$.

Proof. Let $\mathcal{L}(X, \succsim)$ be the set of all sensually completed extensions of $\succsim$. Then, Theorem 1 implies that for the focal consumer, $\mathcal{L}(X, \succsim) \neq \emptyset$. Now, it suffices to show

$$\succsim = \bigcap_{R \in \mathcal{L}(\mathrm{X}, \succsim)} R. \quad (3)$$

Let $\succsim^{\dagger}$ denote the right-hand side of equation (3). Then, the definition $R \in \mathcal{L}(X, \succsim)$ indicates that $\succsim \subseteq \succsim^{\dagger}$. As for the opposite direction $\succsim \supseteq \succsim^{\dagger}$, let $\succsim^{\#} \in \mathcal{L}(X, \succsim)$ be defined as follows:

$$\succsim^{\#} = \succsim \cup \{(x, y) \in X \times X: (y, x) \in (\succsim^{*} - \succsim)\}, \tag{4}$$

where $\succsim^{*}$ is the sensually completed preference of consumptions constructed in the proof of Theorem 1. Hence, for any $x, y \in X$, if x and y are $\succsim$-incomparable, it must be that either $x \succsim^{*} y$ or $x \succsim^{\#} y$, but not both. That is, x and y are ($\succsim^{*} \cap \succsim^{\#}$)- incomparable. So, $\succsim^{*} \cap \succsim^{\#} \supseteq \succsim^{\dagger}$ implies that x and y are $\succsim^{\dagger}$- incomparable. This concludes the proof of $\succsim \supseteq \succsim^{\dagger}$. QED

Theorem 3. For the focal consumer's preference relation $\succsim$ on the set X of all his possible consumptions, there are two sensually completed extensions L_1 and L_2 such that for each pair $x, y \in X$ of $\succsim$-incomparable consumptions, one of the following statements holds true:

- xL_1y and yL_2x;
- yL_1x and xL_2y.

Proof. The conclusion follows for $L_1 = \succsim^{*}$ in equation (7) and $L_2 = \succsim^{\#}$ in equation (4). QED.

By borrowing the concept of order dimensions of posets – partially ordered sets (Dushnik & Miller, 1941), let us similarly define the order dimension of a preference relation $\succsim$ defined on a set X of possible consumptions, denoted $\dim(X, \succsim)$, as the minimum number of sensually completed extensions of $\succsim$ so that their intersection is equal to $\succsim$, if this number is finite, and ∞, otherwise. Symbolically, if $\dim(X, \succsim) \in \mathbb{N}$ (= the set of all natural numbers), then

$$\dim(X, \succsim) = \min\left\{k \in \mathbb{N}: R_i \in \mathcal{L}(\mathrm{X}, \succsim), i = 1, \dots, k, \text{and } \succsim = \bigcap_{i=1}^{k} R_i \right\}.$$

Corollary 1. For each preference relation $\succsim$ on a set X of all consumptions possible for the focal consumer, $\dim(X, \succsim) = 2$.

Proof. The physiological needs of any chosen consumer are multidimensional, meaning that there definitely are consumption choices in X the consumer cannot tell which one of them is preferred over the other. For example, one consumption choice x is from the dimension of housing, and the other y from the dimension of food. Then, no consumer can say that he prefers x over y or y over x, because to survive, he needs both. This fact of real life implies that for each preference relation $\succsim$ on a set X of all consumptions possible for the focal consumer, $\succsim$ has to be incomplete. Hence, this result follows directly from Theorem 3, where $\succsim = \succsim^* \cap \succsim^\#$. QED

5. LINEAR EXTENSIONS OF PARTIAL ORDERS

By slightly modifying the proof of Theorem 1, the following result can be shown.

Theorem 4. Let $\succsim$ be an incomplete partial order defined on a nonempty set X. Then, $\succsim$ can be extended to two linear orders $\succsim^*$ and $\succsim^\dagger$ such that for any pair of $\succsim$-incomparable elements a and b in X, either $(a \succsim^* b) \wedge (b \succsim^\dagger a)$ or $(b \succsim^* a) \wedge (a \succsim^\dagger b)$, but not both, holds true.

The idea behind the argument for Theorem 4 is similar to that of the proof of Theorem 1. In particular, after well ordering X into $\{x^\alpha : \alpha \in I\}$ with $I = |X|$, let x^{β_0} and x^{γ_0} be the elements from N^0, as defined in equation (2), with the least and the second least indexes β_0 and γ_0. Then, we extend $\succsim$ into two partial orders R^0 and S^0 by respectively adding $\left(x^{\beta_0}, x^{\gamma_0}\right)$ and $(x^{\gamma_0}, x^{\beta_0})$, and then taking their individual transitive closures.

Next, repeat this process by singling out x^{β_1} and x^{γ_1} from N^1 the least and the second least indexes β_1 and γ_1 from $N^1 = N^0 - \left\{x^{\beta_0}, x^{\gamma_0}\right\}$. Then, extend R^0 and S^0 to R^1 and S^1 by respectively adding $\left(x^{\beta_1}, x^{\gamma_1}\right)$ and $(x^{\gamma_1}, x^{\beta_1})$, and then taking their individual transitive closures. By applying the transfinite induction on $\alpha \in I$, this process can be completed; and so, the desired linear extensions $\succsim^*$ and $\succsim^\dagger$ of $\succsim$ can be produced as $\succsim^* = \cup R^\varsigma$ and $\succsim^\dagger = \cup S^\varsigma$. For detailed proof of this theorem, please go to the Appendix.

A related result of Szpilrajn (1930) states that for every partial order $\succsim$ on a set S, if a and b are two $\succsim$-incomparable elements of S, there then exist two extensions L_1 and L_2 such that $a\, L_1 b$ and $b\, L_2 a$. In comparison, these extensions L_1 and L_2 are generally a- and b-dependent, while the extensions R^* and S^* in Theorem 4 exist for each given partial order.

A partial order $\succsim$, defined on a nonempty set X, is said to be representable (Ok, 2002), if there are $n \in \mathbb{N}$ (= the set of all natural numbers) and $u: X \to \mathbb{R}^n$, where $\mathbb{R}$ is the set of all real numbers, such that

$$x \succsim y \text{ iff } u(x) \geqq u(y), \text{for any } x, y \in X. \tag{5}$$

To emphasize the n-dimensional Euclidean space $\mathbb{R}^n$, we will say that $\succsim$ is n-dimensional representable. So, as a consequence of Theorem 4, we have the following.

Theorem 5. Let $\succsim$ be an incomplete partial order defined on a nonempty set X. If X is countable, then $\succsim$ is 2-dimensional representable.

Proof. From Theorem 4 or equation (11), it follows that $\succsim$ has two linear extensions $\succsim^*$ and $\succsim^\dagger$ such that $\succsim = \succsim^* \cap \succsim^\dagger$. Because both $(X, \succsim^*)$ and $(X, \succsim^\dagger)$ are countable and linearly ordered sets, there are numerical representations $u_i: X \to \mathbb{R}$ such that for any $x, y \in X$, $x \succsim y$ iff $u_i(x) \succsim u_i(y)$, for $i = *, \dagger$. Therefore, $u = (u_*, u_\dagger): X \to \mathbb{R}^2$ is a 2-dimensional representation of $(X, \succsim)$. QED

This theorem generalizes a conclusion established by Ok (2002), where Ok proves that if X is countable, $\succsim$ is a partial order on X and $\dim(X, \succsim) < \infty$, then $\succsim$ is n-dimensional representable, for some $n \in \mathbb{N}$. In comparison, Theorem 5 avoids the condition of $\dim(X, \succsim) < \infty$, while the uncertain $n \in \mathbb{N}$ is specified to be 2.

Because Theorem 5 deals with representations of binary relations, this is a good place for us to quickly survey some of the excellent works to round up our presentation here. Evidently, due to the lack of transitivity, a preference relation is generally neither representable nor multi-utility representable, as respectively noted by Ok (2002), Evren & Ok (2011), and Bosi & Herden (2012). As for the question of how to represent a general preference relation that is incomplete and nontransitive, Nishimura and Ok (2016) provide an excellent solution.

6. A FEW FINAL WORDS

In this paper, we examine whether or not a preference relation that is incomplete and nontransitive can be extended to a sensually complete and nontransitive binary relation, similar to the case of the extension theorem for partial orders (Szpilrajn, 1930) and that for preorders (Hansson, 1968).

After affirmatively answered this question, we strengthen a result developed by Szpilrajn (1930) greatly through showing that the order dimension of each incomplete partial order and each preference relation is equal to 2. And then, a result by Ok (2002) is refined to that each countable, incomplete and partially ordered set $(X, \succsim)$ is representable by a function $u: X \to \mathbb{R}^2$.

Beyond what are mentioned above, this work establishes the conclusion that each preference relation is equal to the intersection of all sensually completed preference extensions. That generalizes the result of Dushnik and Miller (1941) for partial orders and that of Donaldson and Weymark (1998) for preorders.

The key for us to successfully develop our conclusions with technical arguments lies on the employment of well-ordering of any nonempty set, while the previous works used different techniques, such as the Zorn's lemma. Although these techniques are different, they are all equivalent to the Axiom of Choice. Although this axiom is employed indirectly and differently by each of these involved authors, our application enables us to derive stronger results.

ENDNOTES

1. To see the difference between actual and potential infinities, let us quote the famous vase puzzle (Lin, 1999). Given a vase and a pile of pieces of paper, each of which is labeled with a unique natural number 1, 2, 3, … For step 1, place pieces of paper labelled 1 – 10 into the vase and then take the piece labeled 1

out. For step n, place pieces of paper labeled $10n-9$ to $10n$ into the vase and then take the piece with label n out. For this recursive process, if we think it can be completed, we then look at an actual infinity. And after the process completes, there is no piece of paper left in the vase. On the other hand, if we think this recursive process cannot be completed, then we deal with a potential infinity. In this case, as the process continues without an end, the vase will contain an increasing number of pieces of paper, approaching ∞ as the process continues indefinitely.

2. The symbol $|X|$ stands for the cardinality of set X. In axiomatic set theory (Kunen, 1980), each ordinal number α is equal to the set of all ordinals < α. For instance, 1 = $\{0\}$, 2 = $\{0,1\}$, …, $\alpha = \{\rho: \rho\ (< \alpha) \text{ is an ordinal number}\}$. Within this framework, $|X|$ is simply the least element in $\{\rho: \rho \text{ is an ordinal number and equipollent to } X\}$. That is why we can employ $|X|$ as an index set.

REFERENCES

Alonso, S., E. Herrera-Viedma, F. Chiclana, and F. Herrera. 2010. “A web-based consensus support system for group decision making problems and incomplete preferences.” *Information Science*, 180(23): 4477-4495.

Aumann, R. 1962. “Utility Theory without the Completeness Axiom.” *Econometrica*, 30: 445-462.

Birnbaum, M.H., and R.J. Gutierrez. 2007. “Testing for intransitivity of preferences predicted by a lexicographic semi-order.” *Organizational Behavior and Human Decision Process*, 104: 96-112; doi:10.1016/j.obhdp.2007.02.001.

Bosi, G., and G. Herden. 2012. “Continuous multi-utility representations of preorders.” *Journal of Mathematical Economics*, 48: 212-218.

Cettolin, E., and A. Riedl. 2019. “Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization.” *Journal of Economic Theory*, 181(May): 547-585.

Debreu, G. 1959. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. New Haven and London: Yale University Press.

Donaldson, D., and J.A. Weymark. 1998. “A quasiordering is the intersection of orderings.” *Journal of Economic Theory*, 78: 382-397.

Dubra, J., and E.A. Ok. 2002. “A model of procedural decision making in the presence of risk.” *International Economic Review*, 43(4): 1053-1080.

Dushnik, B., and E.W. Miller. 1941. “Partially ordered sets.” *American Journal of Mathematics*, 63: 600-610.

Evren, O., and E.A. Ok. 2011. “On the multi-utility representation of preference relations.” *Journal of Mathematical Economics*, 47: 554-563.

Fahimnia, B., J. Sarkis, and A. Eshragh. 2015. “A tradeoff model for green supply chain planning: a leanness-versus-greenness analysis.” *Omega*, 54(1):173-90.

Forrest, J.YL. 2013. *A Systemic Perspective on Cognition and Mathematics*. Balkema, The Netherlands: CRC Press, an imprint of Taylor and Francis.

Forrest, J.YL., Y. Liu, T.A. Wajda, E. Köse, and A.A. Oğuzhan. 2019. “How challenges and successes of a dirm depend on its input-output ecosystem." *Proceedings of the 2019 Annual Conference of Decision Sciences Institute*, New Orleans, LA, November 23 – 25, 2019, pp. 2381-2402.

Glasser, W. 1999. *Choice Theory: A New Psychology of Personal Freedom.* Harper Perennial.

Hansson, B. 1968. "Choice structures and preference relations." *Synthese*, 18: 443-458.

Hong, Z.F., and X.L. Guo. 2019. "Green product supply chain contracts considering environmental responsibilities." *Omega*, 83(1): 155-166.

Kunen, K. 1980. *Set Theory: An Introduction to Independence Proofs*. Amsterdam: North-Holland.

Levin, J., and P. Milgrom. 2004a. Consumer theory. https://web.stanford.edu/~jdlevin/Econ%20202/Consumer%20Theory.pdf, accessed February 07, 2022

Levin, J., and P. Milgrom. 2004b. Introduction to Choice Theory. https://web.stanford.edu/~jdlevin/Econ%20202/Choice%20Theory.pdf, accessed September 29, 2022.

Lin, Y. 1999. *General Systems Theory: A Mathematical Approach*. New York: Plenum and Kluwer Academic Publishers.

Lin, Y. 2002. "Some recent progresses in applied mathematics (I)." *Journal of Pure and Applied Mathematics*, 18(2): 99-104.

Lin, Y. (guest editor) 2008. Systematic Studies: The Infinity Problem in Modern Mathematics. *Kybernetes: The International Journal of Cybernetics, Systems and Management Sciences*, 37(3-4): 385-542.

Lin, Y., and B. Forrest. 2011. *Systemic Structure behind Human Organizations: From Civilizations to Individuals*, New York: Springer.

Liu, Y., B.T. Quan, Q. Xu, J.YL. Forrest. 2018. "Corporate social responsibility and decision analysis in a supply chain through government subsidy." *Journal of Cleaner Production,* 208: 436-447.

Mandler, M. 1999. "Incomplete Preferences and Rational Intransitivity of Choice." mimeo, Harvard University.

Mas-Colell, A., M.D. Whinston, and J.R. Green. 1995. *Microeconomic Theory*. New York, NY.: Oxford University Press.

Meng, F.Y., and X.H. Chen. 2015. "A new method for group decision making with incomplete fuzzy preference relations." *Knowledge-Based Systems*, 73(January): 111-123.

Miller, N.H. 2006. *Notes on Microeconomic Theory*. https://nmiller.web.illinois.edu/notes.html, accessed July 11, 2022.

Nishimura, H., and E.A. Ok. 2016. "Utility representation of an incomplete and nontransitive preference relation." *Journal of Economic Theory*, 166 (November): 164-185.

Ok, E.A. 2002. "Utility Representation of an Incomplete Preference Relation." *Journal of Economic Theory*, 104(2): 429-449.

Speer, T. 2007. "A short study of Alexandroff spaces." arXiv:0708.2136 [math.GN].

Suzumura, K. 1976. Remarks on the theory of collective choice, *Economica*, 43: 381-390.

Szpilrajn, E. 1930. “Sur l’Extension de l’Orde Partiel.” *Fundamenta Mathematicae*, 16: 386—389.
Tversky, A. 1969. “Intransitivity of preferences.” *Psychological Review*, 76(1): 31-48.
Vickers, S. 1989. *Topology via Logic*. Cambridge, UK.: Cambridge University Press.

APPENDIX

Proof of Theorem 1. By using the axiom of choice, let us well order X as follows:

$$x^0, x^1, x^2, \dots, x^\alpha, \dots \tag{6}$$

where α is an ordinal-number index from the index set $I = |X|$, so that each nonempty subset $Y \subseteq X$ contains such an element that its ordinal index, as it appears in the list in equation (6), is the smallest compared to those of other elements in Y.

In the rest of this argument, we apply transfinite induction on $\alpha \in I$.

Step 1. Check α values from 0, 1, …, until such least index β_0 in equation (6) that there is $y = x^\beta \in X$ satisfying that x^{β_0} and y are $\gtrsim$-incomparable and $\beta_0 < \beta$. Let us then sensually extend $\gtrsim$ to $\gtrsim^0$ by adding all these $\gtrsim$-incomparable pairs (x^{β_0}, y). Symbolically, we have

$$\gtrsim^0 = \gtrsim \cup \{(x^{\beta_0}, y): y \in X \text{ and } \neg(x^{\beta_0} \gtrsim y) \wedge \neg(y \gtrsim x^{\beta_0})\}.$$

Step 2. Continue checking α, starting from $\beta_0 + 1$, until we find consumption x^{β_1} with the least index $\beta_1 > \beta_0$ in equation (6) so that there is $y = x^\beta \in X$ such that x^{β_1} and y are $\gtrsim^0$-incomparable and $\beta_1 < \beta$. We now sensually extend $\gtrsim^0$ to $\gtrsim^1$ by adding all such $\gtrsim^0$-incomparable pairs (x^{β_1}, y). Symbolically, we have

$$\gtrsim^1 = \gtrsim^0 \cup \{(x^{\beta_1}, y): y \in X \text{ and } \neg(x^{\beta_1} \gtrsim y) \wedge \neg(y \gtrsim x^{\beta_1})\}.$$

Step $\kappa \in I$. Assume that for any ordinal $\tau < \kappa$, three sequences $\{\beta_\rho : \rho \in I_\tau\}$, $\{x^{\beta_\rho} : \rho \in I_\tau\}$ and $\{\gtrsim^\rho : \rho \in I_\tau\}$ have been constructed, for some ordinal $I_\tau \subset \kappa$, so that for any $\rho, \sigma \in I_\tau$, $\rho < \sigma$ implies $\beta_\rho < \beta_\sigma$, the index of x^{β_σ} is the least satisfying $\beta_\sigma > \beta_\rho$ in equation (6) so that there is $y = x^\beta \in X$ such that x^{β_σ} and y are $\gtrsim^\rho$-incomparable and $\beta_\sigma < \beta$, and

$$\gtrsim^\sigma = \bigcup_{\varsigma<\sigma} \gtrsim^\varsigma \bigcup \{(x^{\beta_\sigma}, y): y \in X \text{ and } \neg(x^{\beta_\sigma} \gtrsim y) \wedge \neg(y \gtrsim x^{\beta_\sigma})\}.$$

Find $x^{\beta_{I_\tau+1}}$ with the least index $\beta_{I_\tau+1} > \beta_\tau$, if such a consumption $x^{\beta_{I_\tau+1}}$ still exists, for any $\tau < \kappa$, in equation (6) so that there is $y = x^\beta \in X$ such that $x^{\beta_{I_\tau+1}}$ and y are $\succsim^\tau$-incomparable and $\beta_{I_\tau+1} < \beta$. Let us sensually extend $\{\succsim^\tau : \tau \in I_\tau\}$ to $\succsim^{I_\tau+1}$, where

$$\succsim^{I_\tau+1} = \bigcup_{\tau \in I_\tau+1} \succsim^\tau \bigcup \{(x^{\beta_{I_\tau+1}}, y) : y \in X \text{ and } \neg(x^{\beta_{I_\tau+1}} \succsim y) \wedge \neg(y \succsim x^{\beta_{I_\tau+1}})\}.$$

If the aforementioned consumption $x^{\beta_{I_\tau+1}}$ no longer exists, then the desired sensual extension of $\succsim$ is simply given by $\succsim^* = \cup_{\tau \in I_\tau} \succsim^\tau$.

By using transfinite induction, this constructive process can be completed by checking through all α in I. Let I^* denote the index set $\{\beta_0, \beta_1, \dots, \beta_\kappa, \dots\}$, then the desired sensual extension of the preference relation $\succsim$ is equal to

$$\succsim^* = \bigcup_{\kappa \in I^*} \succsim^\kappa. \tag{7}$$

The consumption set X can be either finite or infinite (for the most likely case of an infinite X, see Debrua, 1959). If X is infinite, the construction of $\succsim^*$ in general involves an infinite process. That is when the assumption that each infinity can be actually (not potentially) achieved is needed to successfully exhaust the list in equation (6) of the consumptions in X and to complete the construction of $\succsim^*$. QED

Proof of Theorem 4. Similar to equation (6), we well order X as follows:

$$x^0, x^1, x^2, \dots, x^\alpha, \dots \tag{8}$$

where $\alpha \in I = |X|$ is an ordinal-number index from the index set I. Next, we apply transfinite induction on $\alpha \in I$.

Step 1. Check α values from 0, 1, …, until such least index β_0 in equation (8) that there is $y = x^{\gamma_0} \in X$ with the least index $\gamma_0 > \beta_0$ so that x^{β_0} and y are $\succsim$-incomparable. Let us then extend $\succsim$ into R^0 and S^0 respectively by adding this $\succsim$-incomparable pair (x^{β_0}, y) and then taking transitive closure and by adding pair (y, x^{β_0}) and then taking transitive closure. Symbolically, R^0 and S^0 are given as follows (Szpilrajn, 1930):

$$R^0 = \succsim \cup \{(x^{\beta_0}, x^{\gamma_0})\} \cup \{(p, q) \in X^2 : p \succsim x^{\beta_0} \wedge x^{\gamma_0} \succsim q\}$$

and

$$S^0 = \succsim \cup \left\{\left(x^{\gamma_0}, x^{\beta_0}\right)\right\} \cup \left\{(q,p) \in X^2 : q \succsim x^{\gamma_0} \wedge x^{\beta_0} \succsim p\right\}$$

so that both $R^0$and S^0 are individually partial orders, $R^0 \cap S^0 = \succsim$ and x^{β_0} and x^{γ_0} are both R^0- and S^0-comparable.

Step $\kappa < |X|$. Assume that for any ordinal $\tau < \kappa$, four sequences $\{x^{\beta_\rho} : \rho \in I_\tau\}$, $\{x^{\gamma_\rho} : \rho \in I_\tau\}$, $\{R^\rho : \rho \in I_\tau\}$ and $\{S^\rho : \rho \in I_\tau\}$ have been constructed, for some ordinal $I_\tau \subset \kappa$, so that for any $\rho, \sigma \in I_\tau$,

- $\beta_\rho < \gamma_\rho$,
- $\rho < \sigma$ implies $\beta_\rho < \beta_\sigma$ and $\gamma_\rho < \gamma_\sigma$,
- the index of x^{β_σ} is the least satisfying $\beta_\sigma > \beta_\rho$ in equation (8) so that there is $y = x^{\gamma_\sigma} \in X$ with the least index such that $\gamma_\sigma > \gamma_\rho$ and x^{β_σ} and y are R^ρ- and S^ρ-incomparable, for $\rho < \sigma$, and
- $$R^\sigma = \bigcup_{\varsigma<\sigma} R^\varsigma \cup \left\{\left(x^{\beta_\sigma}, x^{\gamma_\sigma}\right)\right\} \cup \left\{(p,q) \in X^2 : p \succsim x^{\beta_\sigma} \wedge x^{\gamma_\sigma} \succsim q\right\}.$$

  and

  $$S^\sigma = \bigcup_{\varsigma<\sigma} S^\varsigma \cup \left\{\left(x^{\gamma_\sigma}, x^{\beta_\sigma}\right)\right\} \cup \left\{(q,p) \in X^2 : q \succsim x^{\gamma_\sigma} \wedge x^{\beta_\sigma} \succsim q\right\}$$

  satisfying that both R^σ and S^σ are partial orders, $R^\sigma \cap S^\sigma = \succsim$ and x^{β_σ} and x^{γ_σ} are both R^σ- and S^σ-comparable.

Find $x^{\beta_{I_\tau+1}}$ with the least index $\beta_{I_\tau+1} > \beta_\tau$, if such a consumption $x^{\beta_{I_\tau+1}}$ still exists, for each $\tau < I_\tau + 1$, in equation (8) so that there is $y = x^{\gamma_{I_\tau+1}} \in X$ with the least index $\gamma_{I_\tau+1}$ such that $x^{\beta_{I_\tau+1}}$ and y are R^τ- and S^τ-incomparable and $\beta_{I_\tau+1} < \gamma_{I_\tau+1}$. Let us extend $\cup\{R^\tau : \tau \in I_\tau\}$ and $\cup\{S^\tau : \tau \in I_\tau\}$ to $R^{I_\tau+1}$ and $S^{I_\tau+1}$, where

$$R^{I_\tau+1} = \bigcup_{\tau \in I_\tau} R^\tau \cup \left\{\left(x^{\beta_{I_\tau+1}}, x^{\gamma_{I_\tau+1}}\right)\right\} \cup \left\{(p,q) \in X^2 : p \succsim x^{\beta_{I_\tau+1}} \wedge x^{\gamma_{I_\tau+1}} \succsim q\right\}. \tag{9}$$

and

$$S^{I_\tau+1} = \bigcup_{\tau \in I_\tau} S^\tau \cup \{(x^{\gamma_{I_\tau+1}}, x^{\beta_{I_\tau+1}})\} \cup \{(q,p) \in X^2 : q \succsim x^{\gamma_{I_\tau+1}} \wedge x^{\beta_{I_\tau+1}} \succsim p\} \tag{10}$$

where the third parts of the union expressions in equations (9) and (10) represent respectively the transitive closures. So, from Szpilrajn (1930), it follows that $R^{I_\tau+1}$ and $S^{I_\tau+1}$ are partial orders. And the constructions given above indicate that $R^{I_\tau+1} \cap S^{I_\tau+1} = \succsim$ and $x^{\beta_{I_\tau+1}}$ and $x^{\gamma_{I_\tau+1}}$ are both R^κ- and S^κ-comparable.

If the aforementioned consumption $x^{\beta_{I_\tau+1}}$ no longer exists in the list in equation (8), then the desired extensions of $\succsim$ are simply given by $\succsim^* = R^* = \bigcup_{\tau \in I_\tau} R^\tau$ and $\succsim^\dagger = S^* = \bigcup_{\tau \in I_\tau} S^\tau$

By using transfinite induction, this constructive process can be completed by checking through all α in I or all the listed elements in equation (8) for least indexed incomparable consumptions $x^{\beta_{I_\tau+1}}$ and $x^{\gamma_{I_\tau+1}}$. Let I^* denote the index set of $\{\beta_0, \beta_1, \dots, \beta_\kappa, \dots\}$ or that of $\{\gamma_0, \gamma_1, \dots, \gamma_\kappa, \dots\}$, then the desired linear extensions $\succsim^*$ and $\succsim^\dagger$ of the partial order $\succsim$ are equal to R^* and S^*, respectively, where

$$R^* = \bigcup_{\kappa \in I^*} R^\kappa \text{ and } S^* = \bigcup_{\kappa \in J^*} S^\kappa \tag{11}$$

such that both R^* and S^* are linear orders, $R^* \cap S^* = \succsim$ and every $\succsim$-incomparable pair x and $y \in X$ are both R^*- and S^*-comparable. QED

The Long Run Impact of Pennsylvania's Fracking Boom on Local Residents' Income

David A. Latzko*

ABSTRACT

This paper presents a case study investigation into the effects of the Pennsylvania fracking boom on the incomes of local residents. Per capita income is 11 to 19 percent higher in counties with a high level of fracking activity compared to a comparable synthetic control without fracking activity. In the long run the component of residents' income most impacted by fracking is royalties received by the landowners of drilling sites. Fracking has had a far more modest, fleeting effect on the economies of rural areas of Pennsylvania with small or moderate numbers of wells drilled.

INTRODUCTION

Advances in hydraulic fracking and horizontal drilling techniques opened large areas of Pennsylvania that overlay the Marcellus Shale to natural gas production after the turn of the century with the first unconventional well in Pennsylvania drilled in 2003 by Range Resources-Appalachia in Mount Pleasant Township, Washington County (Harper, 2008, p. 9). Use of hydraulic fracking led to the well producing gas in 2005. Drilling activity soon accelerated across Pennsylvania, peaking in 2011 (Figure 1). Through the end of 2022, a total of 13,861 unconventional wells have been drilled across Pennsylvania (Pennsylvania Department of Environmental Protection, Office of Oil and Gas Management, 2023).

The fracking boom in rural Pennsylvania incited a rush of inquiries into its impact on local employment and income using many different empirical strategies. Input-output studies estimated large gains in employment from that shale gas extraction (Considine, Watson, Entler, and Sparks, 2009; Considine, Watson, and Blumsack, 2010; Wang, Stares, and Young, 2015). Peer reviewed regression research suggests that the fracking boom has had a much smaller effect on rural Pennsylvania economies. DeLeire, Eliason, and Timmins (2014) found that fracking increased local employment slightly but had little effect on wages. Several researchers considered New York State's fracking moratorium in 2008 (a permanent ban was enacted in 2020) to constitute a "natural experiment" to investigate the impact of shale gas development on counties along the New York/Pennsylvania border (Cosgrove, LaFave, Dissanayake, and Donihue, 2015; Hastings, Heller, and Stephenson, 2017; Komarek 2016, Paredes, Komarek, and Loveridge, 2015), generally finding that fracking has had a small impact on county-level employment and income growth. Wrenn, Kelsey, and Jaenicke (2015) estimated that about half of jobs associated with drilling in high activity counties went to non-county residents. Suchyta and Kelsey (2018) also reported that while

* Business and Economics Division, Penn State York, 1031 Edgecomb Avenue, York, PA 17403, Email: dxl31@psu.edu

fracking led to an increase in local economic activity between 2007 and 2010, most of the benefits went to non-residents. Hardy and Kelsey (2015) extended the analysis to other forms of income potentially impacted by Marcellus Shale development using personal income tax data from the Pennsylvania Department of Revenue. They found that increases in royalties, net profits, and gross compensation between 2007 and 2010 were highest in counties with high drilling activity.

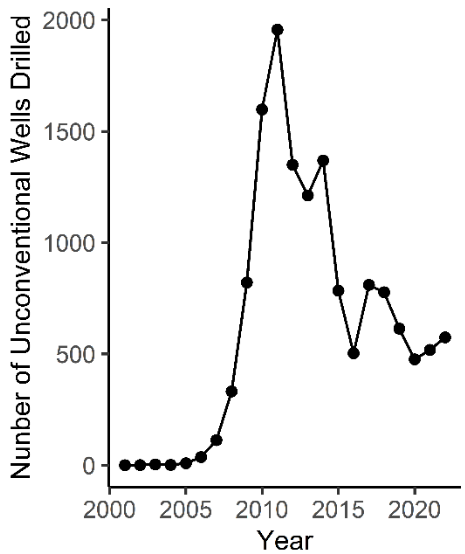


Figure 1. Number of Unconventional Wells Drilled in Pennsylvania by Calendar Year, 2001-2022. Source: Pennsylvania Department of Environmental Protection, Office of Oil and Gas Management (2023).

Hardy and Kelsey (2015) extended the analysis to other forms of income potentially impacted by Marcellus Shale development using personal income tax data from the Pennsylvania Department of Revenue. They found that increases in royalties, net profits, and gross compensation between 2007 and 2010 were highest in counties with high drilling activity.

The purpose of this paper is to assess the long run impact of the Pennsylvania fracking boom on the incomes of local residents. I extend the current literature in three ways. First, the data I utilize covers a much longer time frame, from 1995 to 2020. Existing research stops with 2013, just a few years into the national economic recovery and only two years after drilling activity in Pennsylvania peaked. Since much of the employment associated with natural gas drilling is temporary, lasting as long as wells are being developed, a longer time frame allows for the possibility of disentangling the short run and long run impacts of shale gas development on the incomes of local residents. The longer time frame also permits any differential spatial effects of the business cycle to fade away.

Second, rather than using counties as the units of analysis, I use aggregates of county data to create three regions and examine the effects of fracking on regional incomes. Many of the counties with high levels of drilling activity are economically small. The entry or exit, unrelated to shale gas development, of a single employer can have a large impact on employment and income in that county. Aggregating county data minimizes the effects of this possibility on the results. Also, a regional approach captures some of the geographic spillovers from natural gas development, thereby reducing the possibility of violating the assumption of noninterference necessary for the empirical strategy I adopt.

Third, I use synthetic control methods (Abadie, 2021; Abadie, Diamond, and Hainmueller, 2010; Abadie and Gardeazabal, 2003) to identify the effect of the fracking boom on the incomes of local residents. Rather than supposing that all counties without drilling activity are an appropriate match,I construct synthetic control groups approximately equal in pre-fracking characteristics to the three regional aggregates described below. These synthetic units consist of a weighted combination of several control counties from a pool of Pennsylvania counties in which no unconventional wells have been drilled. The key assumption is that the synthetic control group that is roughly equal to the region with lots of drilling activity in the period prior to the start of unconventional drilling would have, absent the commencement of fracking, been approximately equal to that regional unit in the period after hydraulic fracking began so that income in the synthetic control approximates what income what have been in the regional unit had fracking never occurred. The effect of fracking is estimated by the difference between the income levels in the regional grouping and its synthetic control group in the period after fracking began in Pennsylvania.

DATA AND METHODS

I use annual county-level data for the period 1995-2020. The first unconventional well in Pennsylvania began producing in 2005, giving a pre-fracking period of 10 years and a treatment period of 16 years. My outcome variable is county income per taxpayer. I use the information supplied by the Pennsylvania Department of Revenue (2023) on local residents' taxable income because it is the residents of the communities where drilling is located that most directly bear the costs and inconveniences of that activity. Pennsylvania's personal income tax is levied on wages and salaries, interest and dividends, capital gains, net profits, gambling winnings, and rents, royalties, patents, and copyrights. The data is compiled directly from the tax returns filed with the state and is attributed to the taxpayer's county of residence. Non-county residents, including Pennsylvanians living in other counties, out-of-state workers, and non-resident property owners are not included in the county totals reported by the Pennsylvania Department of Revenue. In the annual summaries published by the Department of Revenue amounts of taxable income by types of income are reported by county along with the number of returns filed. I examine four types of income: (1) total taxable income, (2) gross compensation (henceforth, wages), "income from wages, salaries, tips and other payment for services rendered, before any exclusion for business expenses", (3) business profits, "income, after expenses, from a business, profession, or farm"; and (4) rents, royalties, patents, copyrights (royalties for the rest of the paper), "income received for the use of real or tangible property, the use of a patent or copyright, or the extraction of coal, oil, gas or other minerals" (Pennsylvania Department of Revenue, 2023, pp. 3-4). All county information is based on the county in which the taxpayer resided at the time the return was filed, not necessarily the county in which income was earned. I compute total income, wages, business profits, and royalties per taxpayer for each county. Joint filers are counted as two taxpayers. All values are expressed in 2020 dollars using the CPI-U.

Unconventional wells have been drilled in 39 of the 67 counties in Pennsylvania, but drilling activity is highly concentrated among a handful of rural counties (Figure 2). Six counties, Washington, Susquehanna, Bradford, Greene, Tioga, and Lycoming, account for 70 percent of all the wells drilled in Pennsylvania. I construct three regions, two of which contain counties with high levels of drilling activity. The Southwestern region consists of Greene and Washington Counties. The Northern region contains Tioga, Bradford, Susquehanna, Lycoming, Sullivan, and Wyoming Counties. The latter two counties in this region, while seeing a smaller number of total wells drilled, have a relatively high number of wells per square mile and per capita. The third area, the Central region, is mostly made up of sparsely populated counties with moderate levels of drilling activity: Forest, Elk, Cameron, and Clearfield.

I construct the synthetic controls using the R package “Synth”, specifying that function runs the results over all the possible optimization algorithms and return the result for the best performing method (Abadie, Diamond, and Hainmueller, 2011). I use 15 predictor variables typically associated with income growth potential: (1) 1995-2004 average for population density in persons per square mile, (2) percentage of the county population age 18 to 64 in 2000, (3) percentage of the county population with a bachelor’s degree

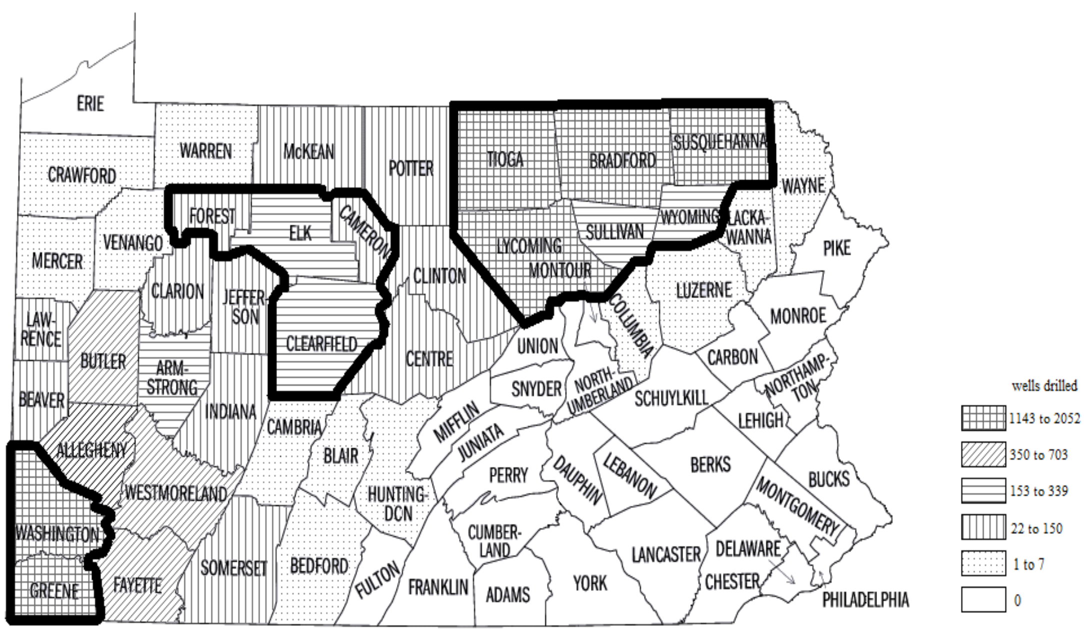


Figure 2. Number of Unconventional Wells Drilled in Pennsylvania by Calendar Year, 2001-2022. Source: Pennsylvania Department of Environmental Protection, Office of Oil and Gas Management (2023).

in 2000, (4) 1995-2004 average for taxable income per taxpayer measured in thousands of 2020 dollars, (5) 1995-2004 averages for nine industrial sector employment shares as a percentage of total employment (construction, farming, finance, government, manufacturing, retail trade, services, transportation, and wholesale trade), and (6) 1989-2004 averages for two industrial sector employment shares as a percentage of total employment (forestry and mining). Age and education levels are taken from the U. S. Census Bureau (2001, pp. 100-101 and 244-245). Industry employment shares were calculated using the data in Table CAEMP25N Total Full-Time and Part-Time Employment by NAICS Industry from the Bureau of Economic Analysis (2022) for 2001 through 2020 and Table CAEMP25S Total Full-Time and Part-Time Employment by SIC Industry for 1989 to 2000. Taxable income per taxpayer is calculated using volumes of Pennsylvania Department of Revenue (2023).

The pool of potential control counties consists of 26 counties in Pennsylvania in which no unconventional natural gas wells were drilled between 2001 and 2022 (see Figure 2). Since fracking occurs in mostly rural areas, I drop Delaware and Philadelphia Counties from the donor pool since both have a population density more than ten times that of Washington County, the county with the densest population included in one of my three regions. I restrict the pool of donor counties to Pennsylvania in order to ensure consistency in the definitions and compilation of the taxable income data as well as in the institutional and governmental arrangements. The potential control counties are geographically distant from the treatment regions making the necessary non-interference assumption more tenable. Of course, to the extent that there are taxpayers such as, for example, a York County resident who owns a hunting cabin in Bradford County and receives royalties from a well on that land, the income gaps reported below will be biased downwards. Table 1 shows the weights of each county in the synthetic versions of the regions for taxable income per taxpayer. These synthetic regions are constructed as the combination of control counties that most resemble the actual regions in terms of pre-unconventional drilling values of taxable income growth predictors. The results are displayed in Table 2, which compares the pre-treatment characteristics of the actual regions with those of the synthetic regions, as well as the unweighted average of the 26 counties in the donor pool. The average of the counties that did not host any unconventional drilling activity does not seem to provide a suitable control group. Prior to 2005, taxable income, population density, and the two human capital variables were substantially higher in the average of the potential control counties than in the three regions.

RESULTS

My estimate of the effect of fracking on local residents' income is the difference between income per taxpayer in the actual set of counties and its synthetic version following the commencement of unconventional drilling in 2005. Figure 3 displays income per taxpayer incomes for the Southwestern region and its synthetic counterpart during the period 1995-2020. Notice that income per taxpayer in the synthetic region very closely tracks income per taxpayer in the actual Southwestern region for the entire pre-fracking period, 1995-2004, averaging a gap of just -$32 per year. This, combined with the high degree of balance on nearly all the income growth predictors between the actual and synthetic regions (Table 2), infers that the synthetic Southwestern region provides a reasonable approximation to the taxable income per capita that would have occurred in the years after 2004 in the absence of unconventional drilling.

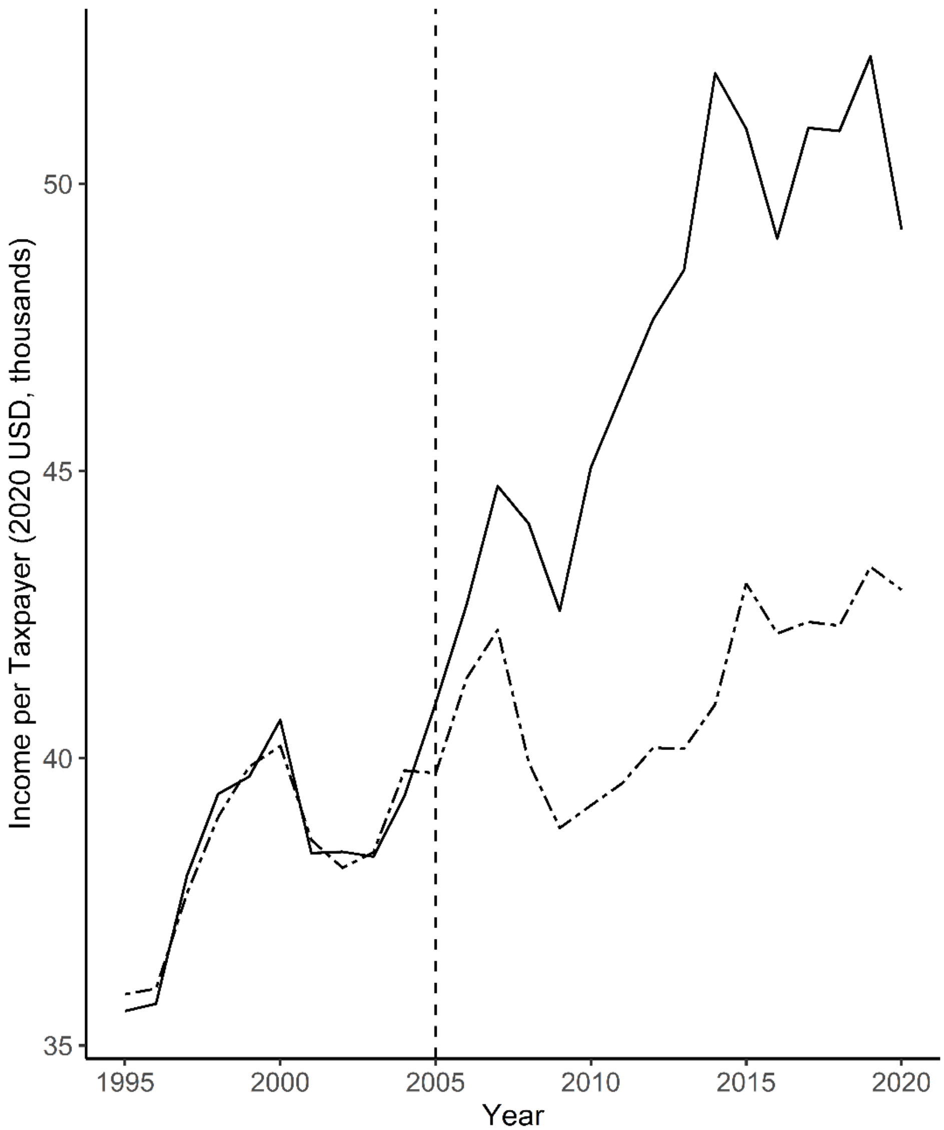


Figure 3. Per-Taxpayer Income in the Southwestern Region (solid line) and its Synthetic Control (double-dashed line).

Figure 3 suggests that fracking had a very large effect on the income of the residents of southwestern Pennsylvania and that this effect increased over time, peaking in 2014, the most active drilling year in Greene and Washington Counties. The magnitude of the estimated impact of fracking in Figure 3 is substantial. The results indicate that for the period 2005-2020 income per taxpayer in these two counties was an average of $6,222 higher each year as a result of fracking; in 2014 income was 27 percent higher in the actual Southwestern region than in its hypothetical no-fracking counterpart. Even in 2020, six years

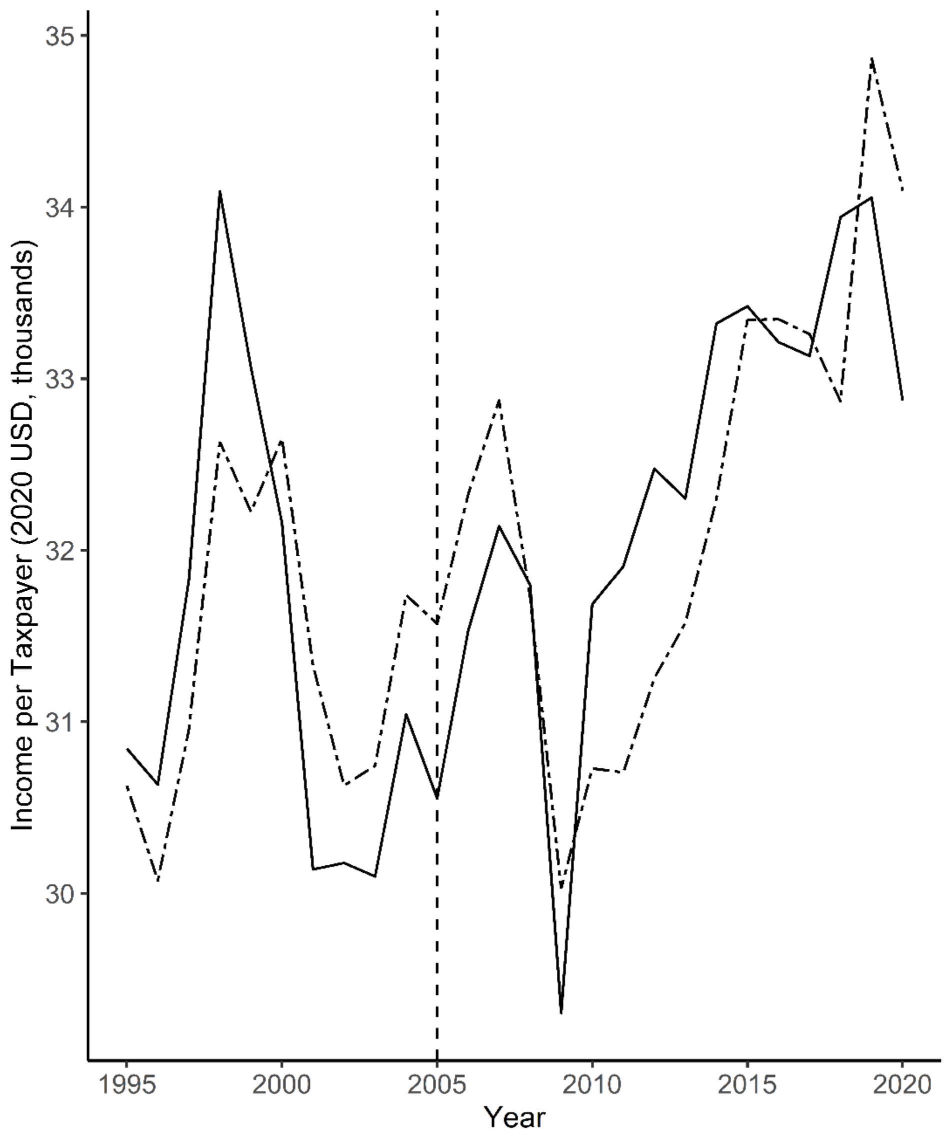


Figure 4. Per-Taxpayer Income in the Central Region (solid line) and its Synthetic Control (double-dashed line).

after the slowdown in drilling activity in this region, income per taxpayer was 15 percent above the synthetic control baseline.

Figure 4 plots the annual income per taxpayer for the Central region and its synthetic control. Unconventional drilling has had only a small impact on the incomes of residents of this area of moderate drilling activity. The income gap between the actual and synthetic region averaged -$49 in the years prior to 2005, again implying a good fit between the actual and synthetic regions, and just $120 per taxpayer a year after 2005. The income gap emerged in 2010, $955 per taxpayer or 3 percent, the year after drilling

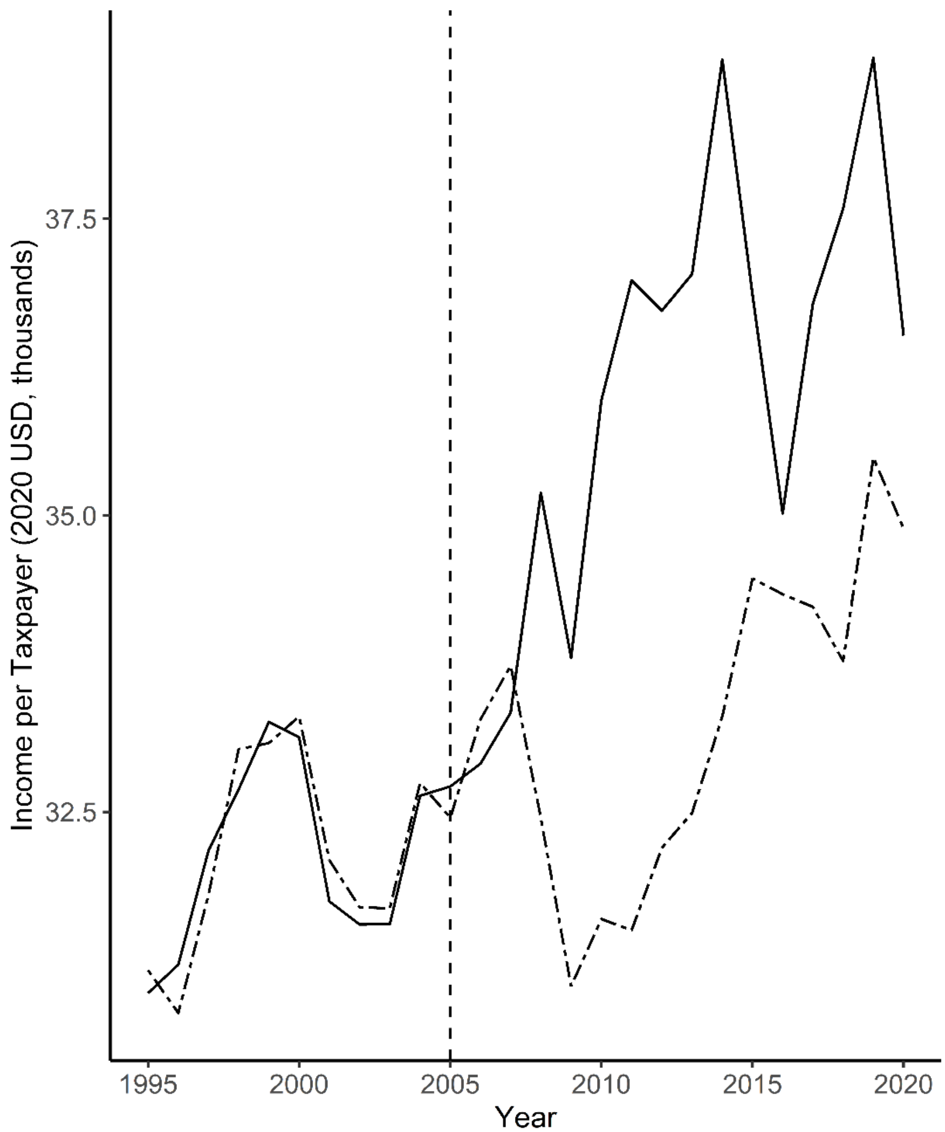


Figure 5. Per-Taxpayer Income in the Northern Region (solid line) and its Synthetic Control (double-dashed line).

began in earnest in Clearfield County. This large gap lasted until 2014, the next to last year of any widespread drilling activity across this four-county region. Just 31 new wells were drilled in this region in 2020 and the income gap was negative $1,223. Income per taxpayer in this four-county region in 2020 was 3.6 percent below the synthetic reference point.

Figure 5 shows the income per taxpayer for the Northern region and the synthetic Northern region. Consistent with these counties experiencing higher levels of drilling activity, the income gaps are again substantial. Since drilling became widespread across northern Pennsylvania in 2008, the gap in taxable

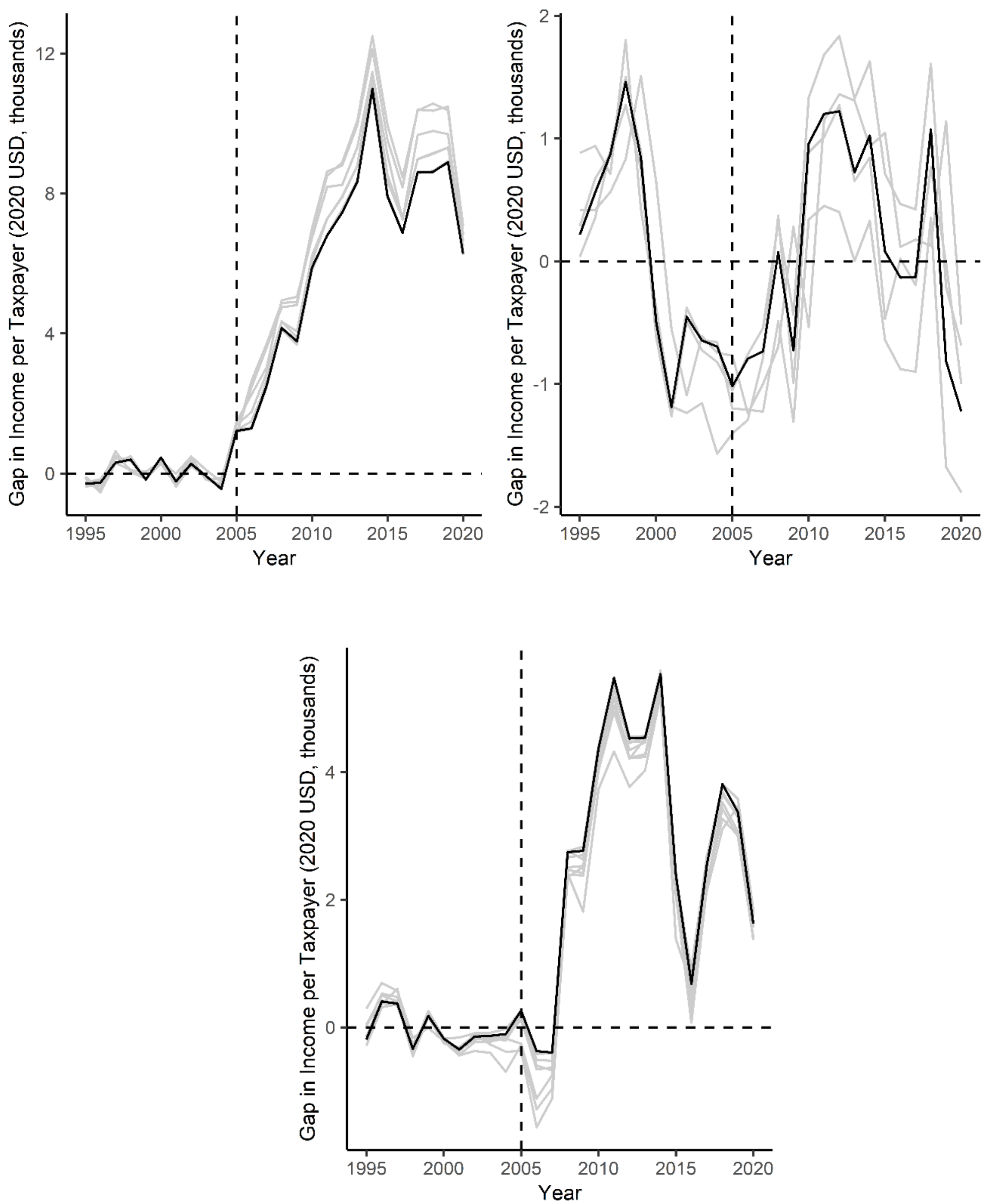


Figure 6. Per-Taxpayer Income Gaps between the Region and its Synthetic Control (black line) and the Region and its Leave-One-Out Synthetic Controls (gray lines) for the Southwestern Region (top left), Central Region (top right), and Northern Region (bottom).

income per capita has averaged $3,415 per year for the Northern region. In 2011, the peak year for well drilling in the region, income per taxpayer was 17 percent greater in the Northern Region compared to its synthetic control. In 2020, relative income per taxpayer was still 4.7 percent higher in the actual set of counties comprising the Northern region.

Figure 6 graphs the results of a robustness check of these results to changes in the county weights in the synthetic controls. I reestimated the baseline model to construct the synthetic regions each time omitting one county in the synthetic control. This procedure allows for an evaluation of the extent to which the results are driven by the most important control counties. The graphs in Figure 6 reproduce the income gaps computed from Figures 3-5 (black lines) while adding the income gaps from the leave-one-out estimates (gray lines). These indicate that the results of the previous analysis for the Northern and Southwestern regions are quite robust to the exclusion of a county from the pool of potential control counties. The leave-one-out synthetic controls show similar income gaps relative to the synthetic controls for both high drilling activity regions. However, several of the leave-one-out models for the Central region, the moderate drilling area with already small income gaps, show much smaller (and sometimes negative) income gaps than the gap from the full model, demonstrating the fragility of the results for the Central region. Fracking may not have had anything more than a small, transitory impact on the incomes of residents of these central Pennsylvania counties.

Figure 7 shows the results of a second robustness test conducted by applying the synthetic control method to each region after reassigning the commencement of fracking to the years 2000 through 2010. The timing of the appearances of the income gaps for the Southwestern and Northern counties is the same for each year. Both these placebo tests and the leave-one-out analyses imply that the measured income gaps are robust to changes in the donor pool of counties and to the assignment of the start of fracking.

To assess the significance of the results I apply the synthetic control method to every county in the donor pool. I reassign the fracking intervention iteratively to one of the 26 control counties, shifting each fracking region to the donor pool. Figure 8 displays the results with the gray lines showing the difference in income per taxpayer for each county and its synthetic control while the black line shows the income gaps estimated above. The figure shows that there is a very low probability of obtaining as large an income gap as the one obtained for the Southwestern region, less than 4 percent. The mean squared prediction error

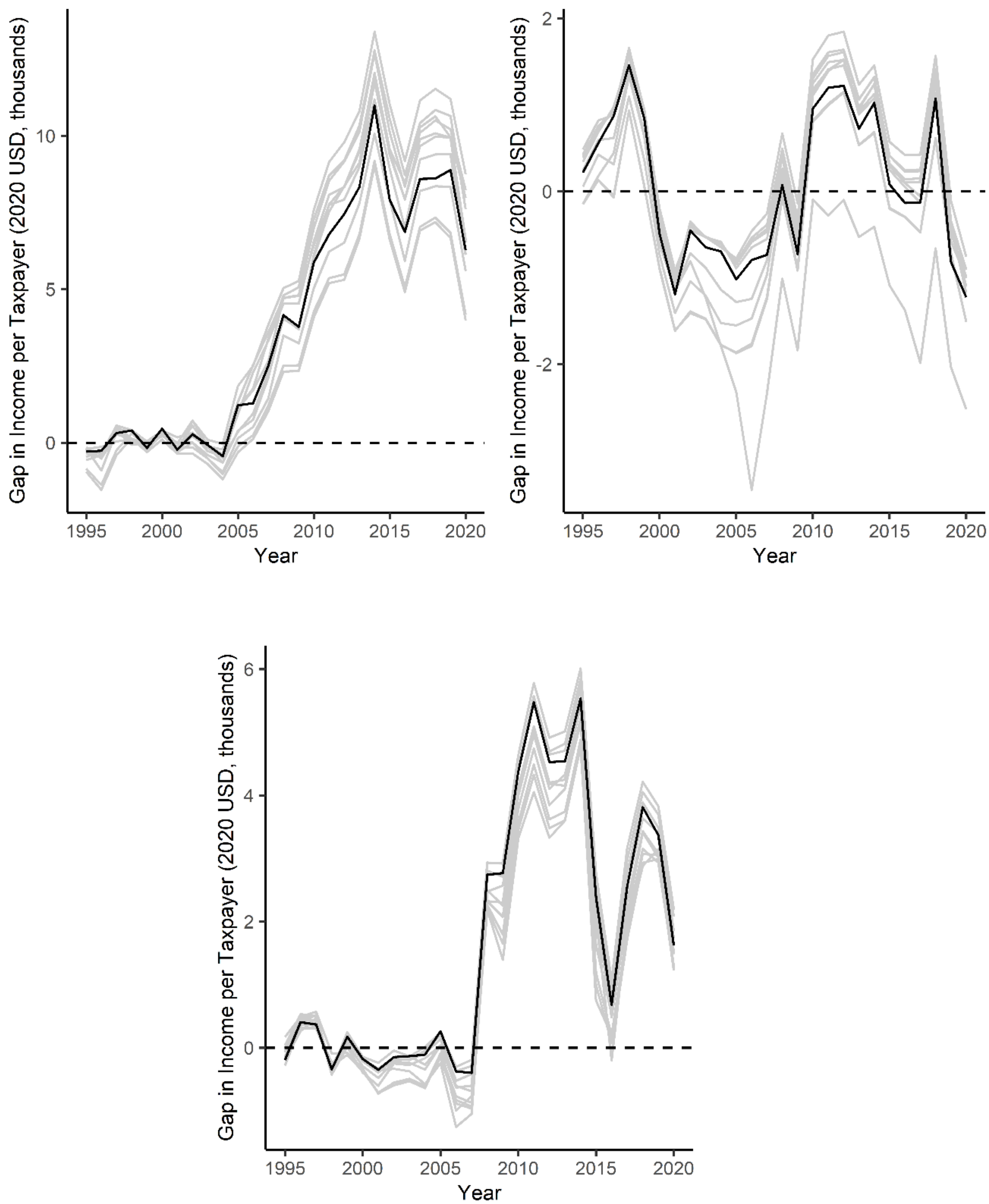


Figure 7. Reassignment Tests of Taxpayer Income Gaps between the Region and its Synthetic Control for 2005 (black line) and 2000-2004 and 2006-2010 (gray lines) for the Southwestern Region (top left), Central Region (top right), and Northern Region (bottom).

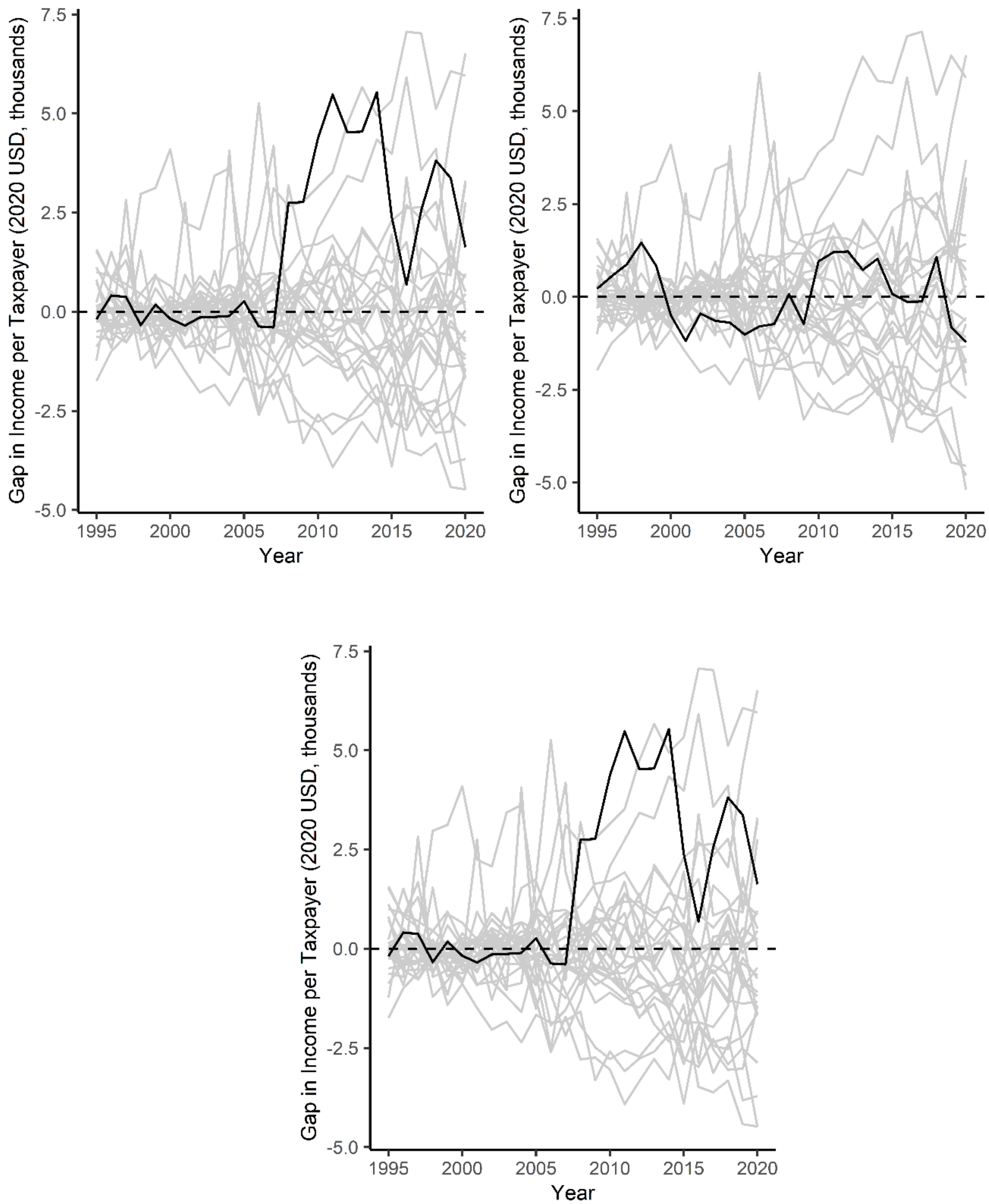


Figure 8. Income Gaps between the Region and its Synthetic Control for 2005 (black line) and the Placebo Gaps in All 26 Control Counties (gray lines) for the Southwestern Region (top left), Central Region (top right), and Northern Region (bottom).

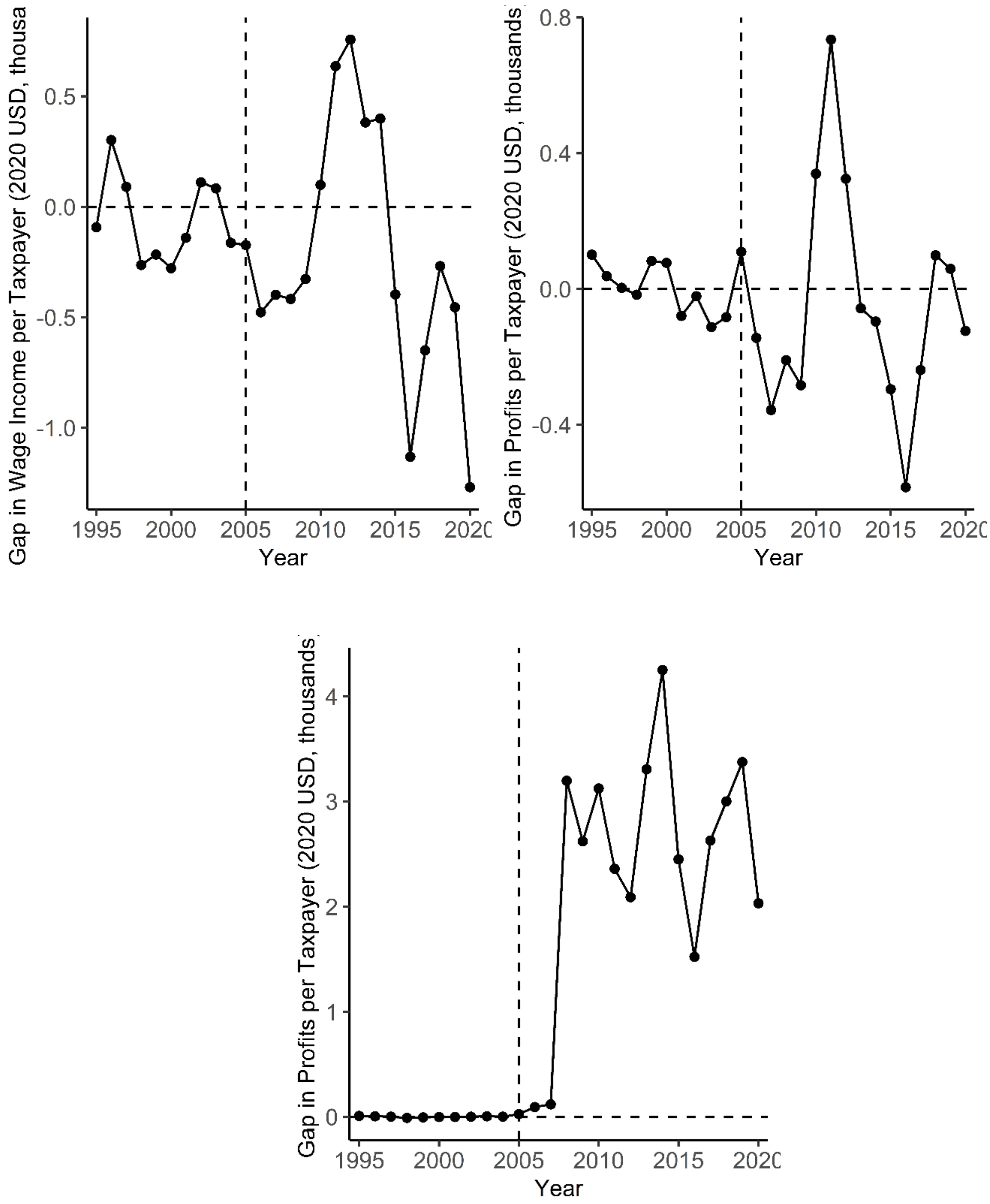


Figure 9. Per-Taxpayer Income Gap between the Northern Region and its Synthetic Controls for Wages (top left), Business

for the region in the post-treatment period is 46.5. The next highest is 26.7. There is about a 7 percent chance of finding larger income gaps for the Northern region. For the Central region, the post/pre-fracking mean squared prediction error is 1.07. If fracking was randomly assigned in the data, the likelihood of obtaining as large a ratio as the Central region's is 92 percent. The income gaps for the Central region cannot be considered statistically significant.

Fracking has had substantial, sustained effects on the incomes of residents in regions of rural Pennsylvania with high levels of drilling activity and a much smaller, possibly nonexistent, impact on the economies of regions with moderate numbers of unconventional wells drilled. I turn the analysis to the types of income affected by fracking. Keep in mind that the Pennsylvania tax returns break down income into several categories. I look at three types of income defined above likely impacted by fracking: wages, net business income, and royalties, constructing synthetic controls for each region for each income type. Figure 9 plots the income gaps between the Northern region counties and its synthetic controls for these three types of income. The graphs show that the long run impact of fracking on the taxable income of residents of this region is entirely due to an increase in royalties income. Typically, landowner agreements with energy companies provide for royalties of a certain percentage of the amount of revenue a well produces. Pennsylvania law guarantees a minimum 12.5 percent royalty rate (Pennsylvania Act Regulating the Terms and Conditions of Certain Leases Regarding Natural Gas and Oil, 1979), although the state's Supreme Court has ruled that deductions for the driller's post-production costs for treating and bringing the gas to market can be taken, even if they result in a royalty below that rate (Kilmer v. Elexco Land Services Company, 2009). The gap in royalties per taxpayer emerged in 2008, the first year of serious drilling across the northern counties. This gap has averaged $2,765 per taxpayer per year since then. The wage income gap was about 2.5 percent above the synthetic control in 2011 and 2012, the peak years for drilling activity in this region. This was also period in which the business profits gap was at its highest, $735 per taxpayer in 2011. Since then, the positive wage and profits gaps have disappeared.

Recall that fracking has had, at best, only a small impact on taxable income in the Central region. Figure 10 shows that most of the observed impact has been on wages. The gap in wage income per taxpayer appeared in 2010 as drilling activity ramped up in Clearfield and Elk Counties and remained persistent through 2019 as new wells continued to be drilled in Elk County, averaging $580 per taxpayer a

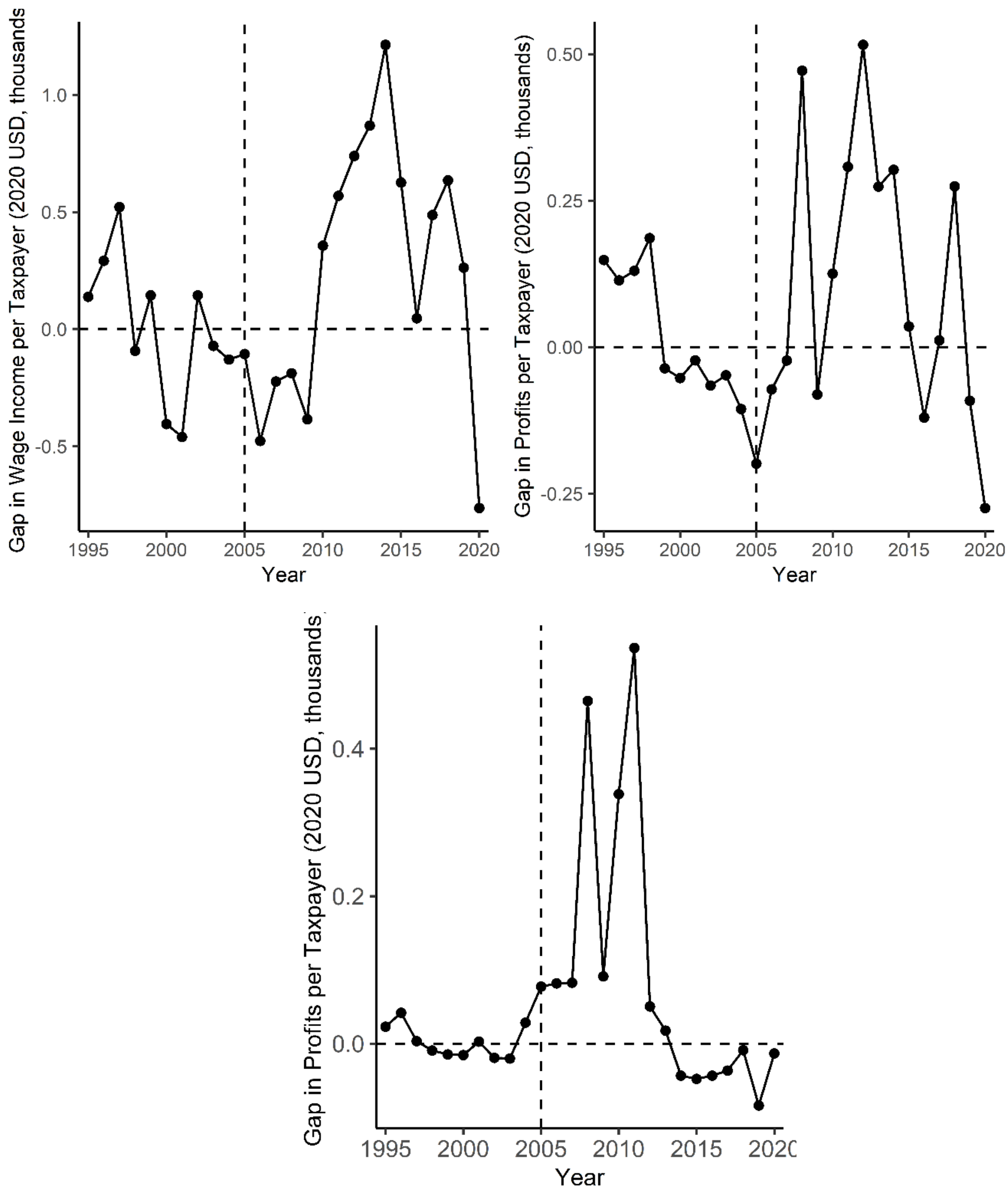


Figure 10. Per-Taxpayer Income Gap between the Central Region and its Synthetic Controls for Wages (top left), Business Profits top (right), and Royalties (bottom).

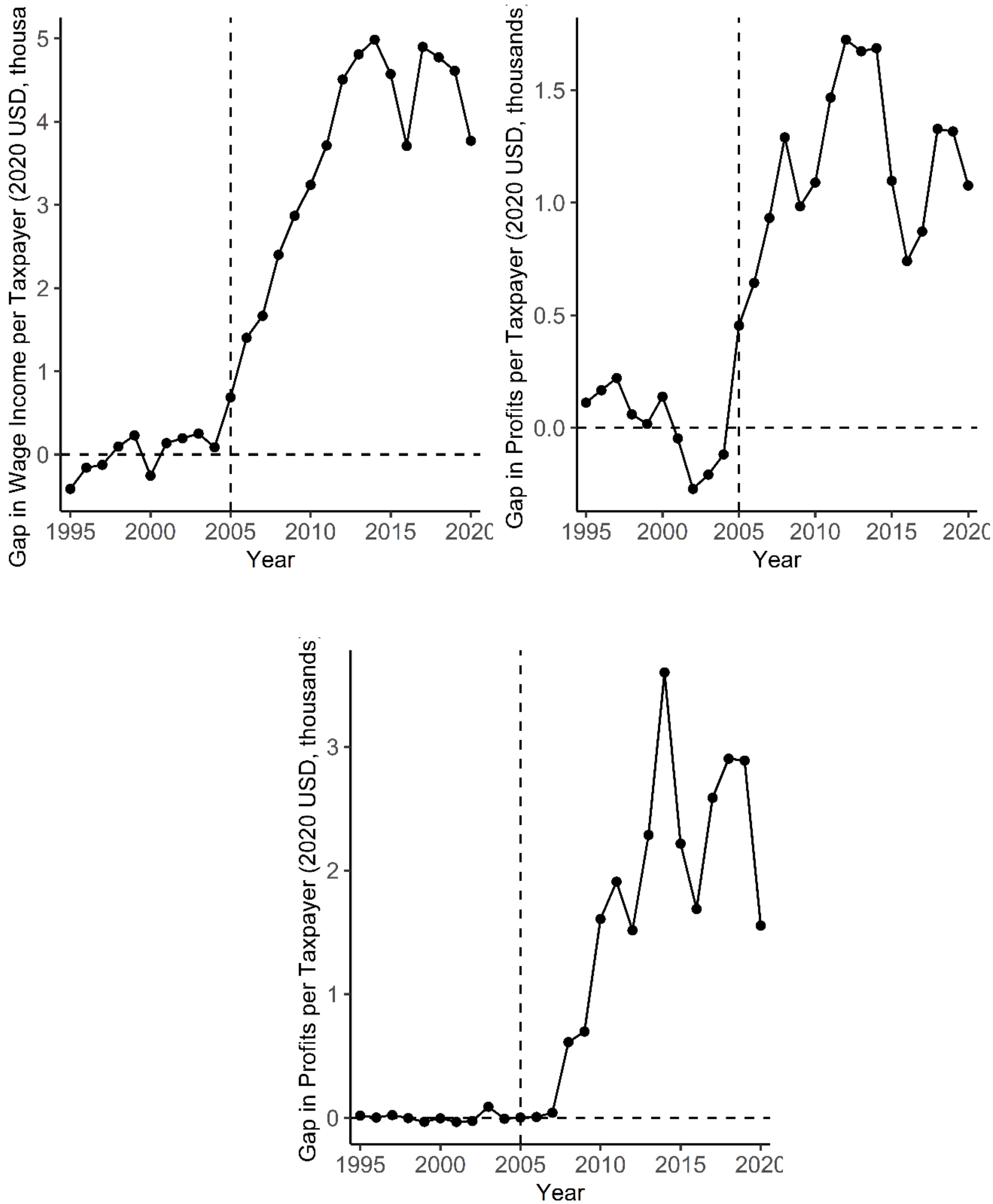


Figure 11. Per-Taxpayer Income Gap between the Southwestern Region and its Synthetic Controls for Wages (top left), Business Profits (top right), and Royalties (bottom).

year. The wage gap, however, turned negative in 2020. Fracking's influence on business profits was smaller and more temporary, essentially vanishing after 2015. Royalties income has been unimportant in the Central region. The gap in royalties income between the Central region and its synthetic control has been about zero every year since 2012.

The substantial consequences of fracking on the incomes of residents of the Southwestern region, two counties with high levels of drilling activity, are mostly the result of higher wages and royalties relative to the synthetic controls. Figure 11 shows the per-taxpayer income gaps between the Southwestern region and its synthetic controls for wages, business profits, and royalties. The wage gap appeared as soon as drilling commenced in southwestern Pennsylvania and has endured since then. Wages per taxpayer were about 11 percent higher, $3,769, compared to the synthetic control in 2020. The annual gap in royalties income has averaged $2,250 since 2010. The gap in net business profits, while also persistent since 2005, has been much smaller, averaging $1,148 per taxpayer per year.

SUMMARY AND CONCLUSIONS

Much research has been conducted into the effects of unconventional drilling on the Pennsylvania economy. However, to date little case study investigation has been produced on its effects on the incomes of local residents. This paper presents evidence of a positive impact on local incomes in areas that experienced relatively high levels of drilling activity. The first part of this study shows a 19 percent annual average gap between income per taxpayer in the two southwestern Pennsylvania counties and the per taxpayer income of a comparable synthetic region without fracking since 2010 and an average income gap of 11 percent a year over the last dozen years for the high-fracking counties across Pennsylvania's Northern Tier. Fracking has had a far more modest effect on local income in rural areas of Pennsylvania with less intensive drilling activity. The income gap per taxpayer for the Central region counties compared to its synthetic alternative averaged about 3 percent during the handful of years, 2010-2014, in which these counties saw moderate numbers of new wells drilled, and the gap, outside of one outlying year, disappeared once new wells almost ceased to be sunk in those counties, meaning that fracking has not had a statistically significant long run effect on local incomes in this area.

The second part of this study establishes that the component of local income most impacted by fracking in the long run in those regions of Pennsylvania with high numbers of wells drilled is royalties. In the northern counties, where the long run income gap is entirely attributable to the gap in royalties income, 5.0 percent of tax returns reported royalties income in 2004, before the start of extensive fracking in Pennsylvania. Of these returns, the average amount was $11,508 in 2020 dollars. Natural gas prices temporarily collapsed in 2020 with the Henry Hub natural gas spot price dropping 39 percent between March 2019 and March 2020 (U. S. Energy Information Administration, 2023), causing Pennsylvania taxpayers to report 6 percent less royalties in 2020 than in 2019. So, for 2019, 17.1 percent of returns from residents of the Northern Region reported an average of $31,372 in royalties per return. A comparable increase in both the percentage of households reporting royalties income and the amount received occurred in southwestern Pennsylvania. The percentage of returns with income from royalties rose from 5.0 percent in 2004 to 15.3 percent in 2019 and the average amount reported more than doubled from $14,739 in 2004 to $31,816 in 2019. Royalties income from fracking is much less important for residents in areas with small or even moderate amounts of drilling activity, with royalties income falling in real terms and making up a smaller share of local income today than prior to the commencement of unconventional drilling in these areas. For the four central Pennsylvania counties making up the Central Region, 5.9 percent of tax returns for 2004 reported royalties income and the average amount was $9,872 in 2020 dollars. In 2019, just 5.4 percent of returns for residents of the Central Region reported an average of $9,453 in income from royalties.

Only a few unconventional wells were drilled into New York's Marcellus Shale before the ban (U.S. Energy Information Administration, 2017), so the total amount of natural gas reserves under New York State is uncertain. The U.S. Geological Survey (2019) estimated undiscovered, technically recoverable natural gas resources for six assessment units in the Marcellus Shale region. Its estimate of mean undiscovered resources for the Western Margin assessment unit, an arc from West Virginia into Ohio and just over the border between New York and Pennsylvania, is one-sixth that for the much smaller geographically Northern Interior unit that encompasses the Northern region analyzed above and one-fifth that of the Eastern Interior which includes the Central region counties. This implies that the long run impact of fracking on the incomes of residents of New York's Southern Tier would likely be quite limited, likely even smaller than this paper's estimate for the Central Region counties.

ENDNOTES

I would like to thank three anonymous reviewers for their especially helpful comments.

REFERENCES

Abadie, A. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59(2): 391-425. DOI 10.1257/jel20191450

Abadie, A., Diamond, A., and Hainmueller, J. 2011. "Synth: An R Package for Control Methods in Comparative Case Studies." *Journal of Statistical Software* 42(13): 1-17. DOI 10.18637/jss.v042.i13

Abadie, A., Diamond, A., and Hainmueller, J. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105(490): 493-505. DOI 10.1198/jasa.2009.ap08746

Abadie, A. and Gardeazabal, J. 2003. "The Economic Costs of Conflict: A Case Study of the Basque County." *American Economic Review,* 93(1): 113-132. DOI *10.1257/000282803321455188*

Bureau of Economic Analysis, 2022. "Regional Data." [Online] Available: https://apps.bea.gov/iTable/index_regional.cfm (July 12, 2022).

Considine, T., Watson, R., Entler, R., and Sparks, J. 2009. "An Emerging Giant: Prospects and Economic Impacts of Developing the Marcellus Shale Natural Gas Play." [Online] Available: https://marcelluscoalition.org/2009/08/n-emerging-giant-prospects-and-economic-impacts-of-developing-the-marcellus-shale-natural-gas-play/ (July 22, 2019).

Considine, T., Watson, R., and Blumsack, S. 2010. "The Economic Impacts of the Pennsylvania Marcellus Shale Natural Gas Play." [Online] Available: An Update. https://marcelluscoalition.org/wpcontent/uploads/2010/05/PA-Marcellus-Updated- Economic-Impacts-5.24.10.3.pdf (July 22 2019)

Cosgrove, B., LaFave, D., Dissanayake, S., and Donihue, M. 2015. "The Economic Impact of Shale Gas Development: A Natural Experiment along the New York/Pennsylvania Border." *Agricultural and Resource Economics Review*, 44(20): 20-39. DOI 10.1017/S1068280500010200

DeLeire, T., Eliason, P., and Timmins, C. 2014, "Measuring the Employment Impacts of Shale Gas Development." [Online] Available: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.642.610 (July 2, 2022)

Hardy, K. and Kelsey, T. 2015. "Local Income Related to Marcellus Shale Activity in Pennsylvania." *Community Development* 46(40), 329-340. DOI 10.1080/15575330.2015.1059351

Harper, J. 2008. "The Marcellus Shale — An Old "New" Gas Reservoir in Pennsylvania." *Pennsylvania Geology*, 38(1): 2-13.

Hastings, K., Heller, L., and Stephenson, E. 2017. "Fracking and Labor Market Conditions: A Comparison of Pennsylvania and New York Border Counties." *Eastern Economic Journal*, 43(4): 649-659. DOI 10.1057/eej.2015.47

"Kilmer v. Elexco Land Services Company, Inc., 990 A.2d 1147, PA" (2010), [Online] Available: http://www.pacourts.us/assets/opinions/Supreme/out/j-78-2009mo.pdf. (May 5, 2020).

Komarek, T. 2016. "Labor Market Dynamics and the Unconventional Natural Gas Boom: Evidence from the Marcellus Region." *Resource and Energy Economics*, 45: 1-17. DOI 10.1016/j.reseneeco.2016.03.004

Parades, D., Komarek, T., and Loveridge, S. 2015. "Income and Employment Effects of Shale Gas Extraction Windfalls: Evidence from the Marcellus Region." *Energy Economics*, 47: 112-220. DOI 10.1016/j.eneco.2014.09.025

"Pennsylvania Act Regulating the Terms and Conditions of Certain Leases Regarding Natural Gas and Oil, General Assembly of the Commonwealth of Pennsylvania, P.L. 183, No. 60." 1979. [Online] Available: https://www.legis.state.pa.us/WU01/LI/LI/US/HTM/1979/0/0060..HTM. (May 5, 2022).

Pennsylvania Department of Environmental Protection, Office of Oil and Gas Management, (2023), "Wells Drilled by County." [Online] Available: http://cedatareporting.pa.gov/Reportserver/Pages/ReportViewer.aspx?/Public/DEP/OG/SSRS/Wells_Drilled_By_County. (January 31, 2023).

Pennsylvania Department of Revenue. 2023. *Personal Income Tax Statistics: Tax Year 2020*. [Online] Available:https://www.revenue.pa.gov/NewsandStatistics/ReportsStats/PIT/Documents/2020_pit_stats.pdf (February 7, 2023).

Suchyta, M. and Kelsey, T. 2018. "Employment and Compensation in the Marcellus Shale Boom: What Stays Local?" *Journal of Rural and Community Development*, 13(4): 87-106.

U. S. Census Bureau. 2001. *County and City Data Book: 2000*. U. S. Government Printing Office, Washington, DC.

U. S. Geological Survey 2019. *Assessment of Undiscovered Gas Resources in the Middle Devonian Marcellus Shale of the Appalachian Basin Province, 2019*. [Online] Available:

https://pubs.usgs.gov/fs/2019/3050/fs20193050.pdf. (February 5, 2023).

U. S. Energy Information Administration. 2017. *Marcellus Shale Play: Geology Review*. [Online] Available: https://www.eia.gov/maps/pdf/MarcellusPlayUpdate_Jan2017.pdf. (February 5, 2013).

U. S. Energy Information Administration. 2023. "Henry Hub Natural Gas Spot Price" [Online] Available: https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm. (February 9, 2023).

Wang, Y., Stares, D., and Young, C. 2015. *An Analysis of the Economic Impact of Unconventional Shale Development in Washington County, Pennsylvania (2011 to 2013)*. [Online] Available: https://issuu.com/washjeff/docs/cepm_report_on_econ_impact_study_20. (July 22 2019).

Wrenn, D., Kelsey, T., and Jaenicke, E. 2015. "Resident vs. Nonresident Employment Associated with Marcellus Shale Development." *Agricultural and Resource Economics Review*, 44(2): 1-19. DOI 10.1017/S1068280500010194

Appendix

Table 1. County Weights in the Synthetic Regions for Taxable Income per Taxpayer.

County	Central	Northern	Southwestern
Adams	0	0	0
Berks	0	0	0
Bucks	0	0	0.225
Carbon	0	0.041	0
Chester	0	0	0
Cumberland	0	0	0
Dauphin	0	0.006	0.027
Erie	0	0	0
Franklin	0	0.003	0
Fulton	0	0.020	0.097
Juniata	0	0	0
Lancaster	0	0	0
Lebanon	0	0	0
Lehigh	0	0	0.069
Mifflin	0.637	0.429	0
Monroe	0	0	0.001
Montgomery	0	0	0
Montour	0.071	0.090	0
Northampton	0	0	0
Northumberland	0	0	0.532
Perry	0	0.175	0
Pike	0.103	0	0.046
Schuylkill	0	0	0
Snyder	0.189	0.236	0
Union	0	0	0
York	0	0	0

Table 2. Taxable Income Growth Predictor Means before Unconventional Drilling. Sources: Author's computations from Bureau of Economic Analysis (2022), Pennsylvania Department of Revenue (2023), and U. S. Census Bureau (2001).

	Southwestern		Central		Northern		
Variables	Real	Synthetic	Real	Synthetic	Real	Synthetic	Average[a]
Income[b]	31.4	31.4	32.1	32.1	38.3	38.3	40.6
Population density[c]	46.5	109.3	194.6	110.4	170.3	413.7	362.5
Adults[d]	59.6	59.0	59.9	60.0	60.4	60.0	60.7
College[e]	8.9	10.4	12.3	10.1	13.2	13.5	15.6
Sectoral shares[f]							
Construction	4.8	4.9	5.4	5.5	7.8	5.8	6.0
Farming	1.3	4.2	4.6	5.1	2.8	2.6	2.8
Finance	4.0	4.7	4.8	4.3	5.1	5.6	6.3
Forestry[g]	0.5	0.8	1.1	1.0	0.7	0.8	0.9
Government	11.4	10.5	12.4	11.7	11.8	11.7	11.9
Manufacturing	22.1	22.1	18.9	20.3	12.0	18.9	16.7
Mining[g]	1.6	0.1	0.6	0.5	3.2	2.3	1.0
Retail trade	16.9	16.4	15.6	16.5	15.6	15.8	15.2
Services	24.7	27.3	25.5	25.7	30.2	30.0	31.2
Transportation	5.1	4.0	4.6	4.8	5.0	4.5	4.4
Wholesale trade	2.5	3.3	3.3	3.1	4.1	3.9	3.4

[a] Unweighted average of 26 control counties
[b] Taxable income per taxpayer in 2020 USD, thousands, average for 1995-2004
[c] Persons per square mile, average for 1995-2004
[d] Percentage of population age 18-64, 2000
[e] Percent of adults with a bachelor's degree, 2000
[f] Percentages of total employment, average for 1995-2004, except where noted
[g] Percentages of total employment, average for 1989-2004

A Theoretical Analysis of Student Evaluations of Instruction Using a Modification of the McKenzie Model

William P. O'Dea*

ABSTRACT

A common belief is that faculty can "buy" higher student evaluations by relaxing their grading standards. Empirical support for this proposition has been mixed. This paper extends McKenzie's model to include two types of students: grade maximizers and knowledge maximizers. Each type of student faces a constraint that shows the rate at which the student can convert leisure time into a better grade or more learning. Instructors can attempt to manipulate this constraint by improving their instructional methods, lowering their grading standards, or reducing the rigor of the course. Analysis of the three strategies shows that the relation between grades and student evaluations can take on any sign, which implies that the relation between grades and student evaluations must be established empirically.

INTRODUCTION

At my college, as at most institutions of higher education, tenure and promotion decisions are based on the faculty member's performance in the areas of teaching, research and service. As a long-serving member of my college's Promotion and Tenure Committee, I have found that evaluating research and service is less problematic than evaluating teaching. In the case of research, our tenure and promotion portfolios contain reprints of articles, which we can read. Journal acceptance rates and impact factors are readily available as are the number of citations for individual articles. Some departments recruit an outside evaluator to assess a candidate's research portfolio. Turning to service, not all service assignments are equally demanding. However, the members of our Promotion and Tenure Committee are experienced senior faculty who are well-informed about which committee assignments are substantive and which are little more than line items on a c.v.

Evaluating teaching is more difficult. In principle, we would like a measure of what students have learned in their courses. We would also like a measure of the extent to which students retain what they have learned. In literature, two approaches have been used to measure student learning. Weinberg et al. (2009) employ a measure of learning that is based on grades in subsequent courses. Finegan and Siegfried (2000) measure learning as the change in the average class score on the third edition of the Test of Understanding College Economics (TUCE) examination between the beginning and end of the semester. These measures of learning were computed as part of one-time research projects. The routine computation of either of these measures of learning as part of our assessment effort would strain the resources of our Institutional Research Office, our faculty and the support staff of our departments.

* Department of Economics, Emeritus, State University of New York College at Oneonta, e-mail: odeawp@oneonta.edu

The only measure of student learning that we have at our disposal is the grade distribution for each course taught by each instructor. Ideally, course grades should reflect the learning that takes place in a course. We hope that in order to earn a higher grade, a student should successfully master more of the material taught in a course. There are circumstances where course grades can be interpreted as a measure of learning with a reasonable degree of confidence. For example, at the United States Air Force Academy, the syllabi for core courses are standardized. The examinations are designed by the team of instructors teaching each course. Each instructor provides a portion of the questions on the examination and is responsible for grading a small part of the examination for the students in all sections of the course. [See: Carrell and West (2010).] Under these circumstances Carrell and West argue that examination grades are an accurate measure of relative student achievement. However, in most of the courses at my institution, examination design and grading are in the hands of the course instructor. Thus, examination grades may reflect the design and rigor of the examinations and the instructor's grading standards as much as what the students actually learn.

At my college, we ask our faculty to provide a portfolio of materials to document their teaching performance. In addition to grade distribution data, we ask our faculty to submit student evaluations of teaching (SET) data along with student free responses. We also ask for course syllabi, course materials such as copies of examinations, peer reviews based on classroom visits, and a self-evaluation of teaching. It has been my experience that in Promotion and Tenure Committee deliberations, the SET data play the dominant role in discussions of faculty teaching effectiveness. More specifically, even though our SET instrument contains 13 items, the bulk of attention is focused on the "bottom line" question which asks the students for an "overall evaluation of teaching effectiveness." This practice is not peculiar to my institution. Becker et al. (2012) report that "the evaluation of teaching at almost all schools still relies heavily and almost exclusively on SETs (p. 332)." Members of my College's Promotion and Tenure Committee are aware of the allegation that instructors can "buy" better SET ratings by relaxing their grading standards. Therefore, Committee members tend to look at SET data in conjunction with grade distribution data. The operating premise is that as long as a faculty member's grade distribution is in line with that of his/her department then the faculty member is not "buying" better SET scores. Thus, the SET scores themselves reflect the faculty member's relative teaching effectiveness and we can then skim over the rest of the teaching portfolio. In fact, we are especially impressed when a faculty member combines strict grading practices with above average SET scores.

The belief that favorable SET ratings can be purchased is due to the fact that grade inflation at American college and universities coincided with the introduction of student evaluations (Nelson and Lynch 1984). In the economics of education literature, there has been a concerted effort to uncover the relationship between grades (actual or expected) and SET scores. In a nutshell, the findings are all over the place and taken as a whole do not provide a ringing endorsement for the proposition that the recipe for good SET scores is easy grading. A number of authors such as DeCanio (1986) and Ragan and Walia (2010) find that the impact of grades on SET ratings is statistically insignificant. Other authors such as Boex (2000), Kelley (1972), Nelson and Lynch (1984) and Zangenzadeh (1988) report a positive and statistically significant relationship between grades and faculty SET ratings. But their estimated coefficients are small and indicate that the payoff to a grade inflation strategy is not worth the trouble. Based on his results, Kelley (1972) calculates that if a professor who gave mostly Cs were instead to give As and Bs in equal proportion his/her SET score would improve by less than .1 of a point. At my college, a .1 of a point improvement in a faculty member's SET rating would be considered pure noise. Other authors such as Dilts (1980), Isely and Singh (2005), and McPherson (2006) find that the impact of grades (or expected grades) on SET scores is somewhat larger. McPherson, for example, reports that that a 1 point increase in the average grade increases the instructor's SET score by .3 of a point. The strongest relations between grades and SET scores that I've encountered in the literature are those reported by Weinberg et al. (2009). In principles of microeconomics, they found that a one point increase in the average grade received by students would increase an instructor's SET rating by between .37 and .48 of a point depending on the model specification. In principles of macroeconomics and intermediate microeconomics, the improvement in the SET score produced by a 1 point grade increase was even larger ranging from .47 to .94 of a point in principles of macroeconomics and from .52 to 1.05 of a point in intermediate microeconomics.

It is important to note that even if SETs are not overly sensitive to faculty grading practices that does not mean that the SET scores are only measuring the students' perceptions of teaching effectiveness. For example, Hammermesh and Parker (2005) found that more attractive instructors receive significantly better SET scores. In some of their model specifications, Weinberg et al. (2009) discovered that female instructors received lower SET ratings than their male counterparts. These are not the sorts of considerations that we would like to influence student evaluations.

Given the failure of the empirical literature to establish a clear connection between grading standards and SET scores, a number of authors beginning with McKenzie (1975) have developed analytical models that investigate the impact of faculty grading and teaching strategies on the SETs. These models start from the premise that students are welfare maximizers, and that welfare depends on course grades and leisure. This paper will extend the McKenzie model by assuming that courses are taken by two types of students: students whose interest is the grade they receive and students who are interested in mastering the subject matter. In a perfect world, this distinction should not matter. Ideally, higher grades should indicate greater mastery of the subject matter of the course. But, given the control of individual faculty members over examination design and grading, this does not have to be the case. I have seen courses where the median grade is an A and the work requirements are either busy work or no work.

LITERATURE SURVEY

McKenzie's (1975) objective was to develop an economic model of students and faculty behavior that would provide theoretical guidance for empirical studies of student evaluation instruments. His model employs a standard setup that is familiar to anyone who has ever taken or taught intermediate microeconomics. He assumes that students are welfare maximizers and that the arguments in their utility functions are leisure and courses grades. The student selects the combination of grades and leisure that maximizes his/her welfare subject to a constraint. In McKenzie's model, the constraint is a transformation curve that shows the student's ability to transform leisure into a better grade in a given course. McKenzie assumes that the marginal rate of transformation between leisure and the course grade is constant. Therefore, the constraint graphs as a straight-line. The position of the transformation curve depends on the student's ability and interest in the course and on the instructor's grading and instructional practices. The student would maximize welfare by selecting the combination of leisure and course grade on the transformation curve that enables him/her to reach the highest indifference curve.

The instructor can influence the student's optimization problem by moving the transformation curve to the right, which would enable the student to reach a higher indifference curve. As is standard in literature, McKenzie assumes that students will reward teachers who increase their welfare with better course evaluations.

McKenzie then explores various methods that instructors can employ to influence the transformations curves. One possibility is to teach the course more effectively, which would enable students to earn a higher grade for any given input of their own time. However, the instructor would probably have to devote more time to the course, which would leave the instructor with less time for other activities such as research and family. He then considers a variety of ways in which instructors can inflate grades. While the methods differ somewhat, they all move the student transformation curves to the right and thus enable the students to reach higher indifference curves.

McKenzie's analysis provides an explanation for the weak relationship between grades and SET scores observed in the literature. To illustrate, assume that a professor shifts the transformation curve to the right in a parallel manner by adding a uniform ten points to each student's test score while maintaining the same cutoffs for an A, B, C, etc. This shift would produce an effect analogous to the income effect. As long as leisure and the course grade are both normal, both the course grade and the amount of leisure would increase. The key point is that the observed class average would increase by less than ten points. In essence, as Lichty et al. (1978) observe, students take part of the benefits of grade inflation in the form of additional leisure.

McKenzie's analysis focuses on one faculty member's decision to inflate grades. Lichty et al. (1978) extend McKenzie's analysis by recognizing that students typically take not just one course but a schedule of courses. In this broader formulation, an individual instructor who inflates grades would still be rewarded with better evaluations. But it is possible that instructors who maintain their standards might also see an improvement in their SET scores. For example, consider a situation where the instructor of a general education class relaxes his/her grading standards. This would free up time that was previously devoted to the course for other uses. One of these "other uses" might be to devote more time to an important course in the major. More time spent studying for this major course should lead to a better course grade and thus a better SET rating for the instructor of this course.

Needham (1978) employs a different setup to model the relation between grading policy and student evaluations. In his model, students allocate their time between a set of courses with the objective of maximizing their grade point averages. In each course, a higher grade requires that the student devote more time to the course. In the optimal allocation, the marginal grade in each course would be equal. The optimal amount of time spent studying would occur where the sum of the marginal grade curves for each course is equal to the marginal opportunity cost of academic effort in terms of the value of the income or leisure foregone. Needham then assumes that students rate their courses on the basis of the grade received per unit of effort expended. Needham's analysis has an interesting implication: "the relationship between the grades a student receives in different courses and the student's relative rating of these courses could be positive or negative (p. 39)." A student would give a lower rating to a course in which s/he earns a higher grade if the student views the effort required to earn the higher grade as excessive. In Needham's model, instructors have two levers to influence their student evaluations. They can relax their grading standards and assign students a higher grade for each level of effort. Alternatively, they can reduce the effort required to earn a given grade by improving their instructional methods or by watering down the class.

All three of these papers arrive at the same conclusion: instructors can "buy" higher SET scores by relaxing their grading standards. They also provide a theoretical basis for the failure of the empirical analyses of the determinants of SET scores to find a consistent relationship between grades and SET ratings. These papers all approach the problem from a student perspective. This raises two questions: what do instructors want and what does the institution want? While instructor motivations are mentioned in these papers, they are not a formal part of the analysis. Instructors can also be viewed as welfare maximizers. The welfare of instructors can be expected to depend on a number of factors such as how much their students learn, their ability to meet their institution's promotion and tenure requirements, and their own leisure time activities. Like the students, they have a limited amount of time at their disposal. Colleges are concerned with meeting their strategic objectives which include teaching, research and service and maintaining their reputations. McKenzie (1975) and Lichty et al. (1978) focus on the behavior of an individual instructor. While one instructor might be able to use grade inflation to bolster his/her SET ratings, the question is: can all instructors pursue the same strategy simultaneously and, if they do, what will happen to the institution's reputation. These papers do discuss these issues but not in a formal manner. Kanagaretnam et al. (2003) and Love and Kotchen (2010) develop models in which the students, instructors and the institution all have a formal role.

Kanagaretnam et al. (2003) assume that student welfare depends on grades and the effort needed to earn these grades. The student selects the amount of learning effort that maximizes expected utility. Kanagaretnam et al. then derive the expression that characterizes the certainty equivalent of the student's expected utility. The student's expected utility depends on native ability, the instructor's teaching effort and grading policy, and class size. Class size, which is under the control of the institution has an inverse impact on student satisfaction. As class size increases, learning becomes more difficult. The student's expected utility equation then becomes a constraint in the instructor's maximization problem because the instructor has to produce a minimum level of satisfaction to ensure that the student evaluations meet the institution's expectations. The instructor's welfare depends on the level of learning of his/her students and the effort devoted to teaching. Kanagaretnam et al. treat maximizing student learning as being equivalent to minimizing grade inflation. The instructor then solves for the level of grade inflation and teaching effort that minimizes the weighted sum of grade inflation and effort disutility subject to the constraint that s/he delivers at least the minimum level of student satisfaction necessary to meet the college's expectations. The upshot is that the relation between student grades and student evaluations depends on the emphasis that the institution places on SET scores. As long as the institution places a low emphasis on SET scores, the instructor will have no incentive to inflate grades. Once the emphasis on SET ratings passes a threshold, the instructor resorts to grade inflation, which results in a positive relation between grades and student satisfaction/evaluations. A consequence of the overemphasis on SET scores is that student effort (and therefore student learning) decreases.

Love and Kotchen (2010) also develop a model in which the students, faculty and institution all play a formal role. In their model, student welfare depends on future earnings, leisure, and the quality of instruction. In turn, future earnings depend on the student's grades, which depend on the amount of time the student devotes to the course and the instructor's grading standards. The student then looks for the level of study time that maximizes welfare subject to a time constraint. Of course, the optimal amount of study time depends on the instructor's grading policy. The instructor's welfare depends on the level of his/her course evaluations, the amount of time devoted to research, and the amount of time students devote to the course. The last consideration implies that the instructor gets a higher level of satisfaction from teaching harder working, more motivated students. The instructor faces a time constraint. The institution affects the instructor's choice through the weights that it assigns to teaching and research in personnel decisions. As a problem-solving strategy, Love and Kotchen insert the students' problem into the instructor's maximization problem in which case the instructor's choice variables become the grading policy and the amount of time spent delivering the course. The institution can then influence the behavior of teachers and students by changing the weights assigned to teaching and research. Love and Kotchen find that as the institution places more emphasis on teaching (as measured by student evaluation scores) faculty respond by loosening their grading standards. The impact on the effort put into the course by students and faculty is ambiguous but both can decrease. One way the institution can avoid these problems is to further constrain faculty behavior by imposing grade targets. For example, the college might specify that the maximum mean grade in a 100 level course is a 3.2. Interestingly, they also find that as the emphasis placed on research increases, the faculty will also respond by inflating their grades.

What all five of these models have in common is that student welfare is a positive function of grades. This raises the question about the importance of what students actually learn in the course. Lichty et al. (1978) simply assume that the "student's grade in any course represents a positive reflection of that student's learning achievement (p. 4)." Love and Kotchen (2010) assume that "students value the educational quality of their classes (p. 151)." But as noted, they measure educational quality by expected future earnings which depend on the grade received. There is evidence that student learning has a statistically insignificant impact on course evaluations. For example, Weinberg et al. (2009) found that learning as measured by student performance in follow up classes has a statistically insignificant impact on student evaluation scores. Carell and West (2010) reached a similar conclusion. They found a negative relationship between teacher evaluations and learning in subsequent courses, which they termed "deep learning."

Even if students care about grades rather than learning, that does not mean that they won't learn. A professor who cares about what his students learn will design his/her course so that learning is a necessary condition for earning a good grade. However, in a course where grading standards are lax and expectations low, student grades would provide a weak indicator of what the students have learned. A student who performs well in a watered down course might have learned relatively more than the other students in the course but not have learned much in any absolute sense.

In the field of labor economics, an unresolved debate is whether the purpose of higher education is to develop the student's human capital or help the student send a signal to prospective employers about such difficult to measure qualities as native ability or perseverance (Kjelland, 2008 and Weiss, 1995)). Given the focus of this paper these two perspectives are not necessarily mutually exclusive. A student might be very interested in constructing a college record that sends a compelling signal to a potential employer. To that end, the student would want to compile a respectable GPA, participate in clubs and athletics, and earn certifications. Given this objective, one of the criteria on which the student might evaluate a course is the grade earned. However, the student might also be interested in what s/he learns in some courses. Consider the case of an accounting major who plans to take the CPA examination. This student would be interested in his/her GPA because accounting firms employ a minimum GPA when screening applicants for initial job interviews or summer internships. This student would likely evaluate general education courses on the basis of the grade s/he receives. However, this same student could also evaluate an upper level accounting course such as auditing or corporate taxation on the basis of the learning that occurs. One of my advisees described one of his upper level accounting courses as "hard, but a good hard." His point was that while the course was difficult and time-consuming, he believed that it would equip him to pass part of the CPA examination.

In this paper, I will develop a variation of the McKenzie model in which each course is taken by two types of students: students who evaluate the instructor's performance solely on the basis of the grade received (type A); and students who base their evaluation solely on the basis of what they learn in the course (type B). While type A students are focused on grades, they will learn if mastering the material in a course is a necessary condition for getting a good grade. Of course, if they could earn an acceptable grade with less work, they wouldn't complain. A type B student in one course could very well be a type A student in other courses. Admittedly, my assumption that there are only two types of students is a bit extreme. A more reasonable assumption is that between the two extremes there is a continuum of students whose assessment of teaching effectiveness is based on both grades and learning (or perceived learning). The weights places on each of these considerations would vary between students. Assuming that there only two types of students simplify the exposition and focuses attention on the impact of differences in student motivation on student evaluation of teaching.

In the literature, there have been other attempts to introduce differences in student traits into the analysis. Most notably, Grimes et al. (2004) differentiate between students with an internal locus of control and students with an external locus of control. Students with an internal locus of control accept responsibility for their performance in a course while students with an external locus believe that their behavior does not determine their performance but that the responsibility for their performance lies in the hands of other parties. They show that students with an external locus of control tend to be harsher in their evaluations of instructor performance. In my model, I will assume that type B students are internally motivated. Type A students can be either internally motivated or externally motivated. Just because a student is grade-focused does not mean that the student views the grade s/he receives in a course as being beyond his/her personal control.

MODEL DESCRIPTION

To simplify the analysis, I will assume that students take one course at a time. The thirty week academic year would be broken down into ten three week segments. This simplifies the students' choice problem into allocating time between leisure and one course rather than allocating time between leisure and a schedule of courses. I also assume that type B students can accurately gauge their level of learning in a course using a system such as Bloom's taxonomy.

My model follows McKenzie's structure with the stipulation that each course is taken by two types of students. I assume that utility is a function of leisure and either the grade received in the course or the level of learning. Each student's indifference curves are convex to the origin. The slope of the indifference curve reflects the willingness of a student to sacrifice leisure for a better grade or more learning. The slope would be influenced by such factors as the student's interest in the subject matter of the course and the student's aptitude for the material. Departments can attempt to give students the incentive to work harder. For example, my department requires students to earn a C or better in each course required for the major. If a course is a prerequisite for a higher level required course, students earning less than a C must repeat the lower level course and earn a C before they can enroll for the higher level course. The need to repeat a course could conceivably delay a student's graduation by a semester. Even at state university tuition, any delay would be costly.

The student is constrained by a transformation curve that shows the student's ability to convert time into grades (or learning). Unlike McKenzie, I will assume that the transformation curve is concave to the origin, which implies that the marginal rate of transformation between leisure and grades (or learning) is increasing. The shape of the transformation curve in part reflects the instructor's expectations regarding the amount of learning required to achieve each grade level. It also reflects the student's study skills. An increasing marginal rate of transformation would result if instructors expect an equal increment of learning to achieve each higher grade level and if time spent studying by the student exhibits diminishing marginal returns like other production activities in the short run.

The position of the transformation curve in the horizontal plane depends on a constellation of factors that are particular to the individual student, the course instructor and the institution. In the case of the student, the key determinants are: the student's ability as measured by the student's SAT/ACT scores and/or high school average, the student's background in the course subject, and the student's locus of control. Kjelland (2008) posits that students with an external locus of control are less motivated and therefore don't perform as well as students with an internal locus of control. The position of the transformation curve is also affected by the effort the instructor puts into the course, how organized the instructor is, and the instructor's test design and grading standards. There is also evidence that as the number of students in a course section increases student performance decreases. [See: Becker and Powers (2001) and Bedard and Kuhn (2008).] The number of students in a section is influenced by the institution's willingness to provide faculty lines to the department. The institution can also influence the position of the transformation curve through the weights it places on teaching and research in its reward structure.

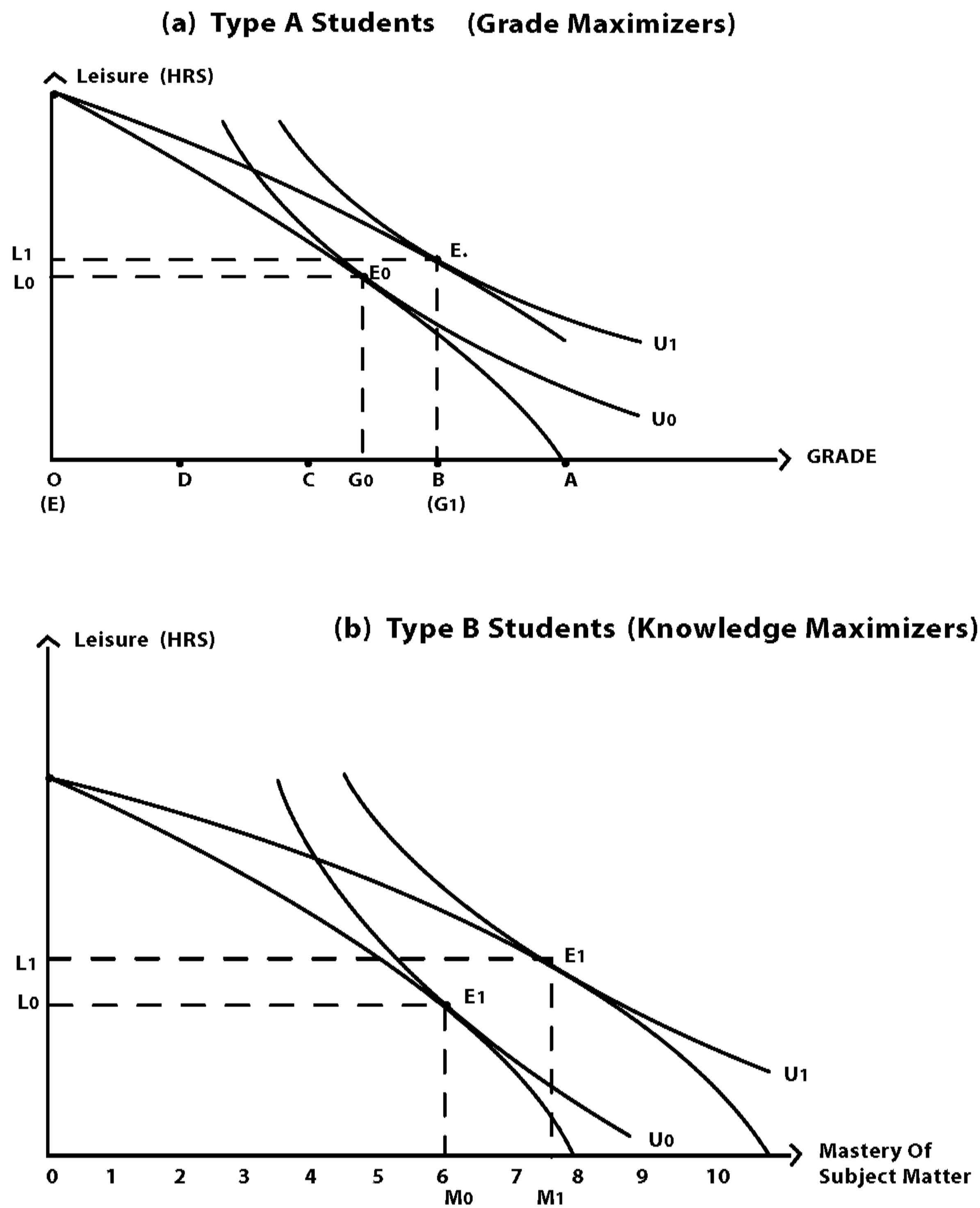
The transformation curve should not be viewed as a deterministic relation in the sense that a given level of time commitment guarantees a particular grade. Rather, the transformation curve reflects the student perception of his or her ability to transform time into either a higher grade or more learning. Ultimately, determination of the course grade is in the hands of the instructor.

The welfare maximizing combination of leisure and learning (or grade) occurs where the transformation is tangent to the highest attainable indifference curve. [The point notated E_0 in figure 1 below.] The welfare maximizing point would correspond to the grade the student expects to receive given the time s/he has committed to the course. Of course, if the student underestimates the difficulty of the course, the actual and expected grades would differ. Since SET forms are (least at my institution) are filled out before the final grade is known, the SET scores would reflect the grades students expect to earn. If a student faces a C or better requirement, I expect that the welfare maximizing point would correspond to (at least) a grade of C. Otherwise put, I find it hard to believe that welfare maximizing students would select a level of time commitment that they expect would require them to repeat the course, especially if they need the course in order to graduate on time.

ANALYSIS

As noted in McKenzie's model, if a faculty member wishes to improve his/her student evaluations, it is necessary to take some action that shifts the student transformation curve to the right, which would enable students to reach higher indifference curves. Again, the premise is that if the students perceive themselves to be better off they will reward the instructor with better evaluations. In this section, we will explore three strategies that a faculty member can employ to shift the transformation curves to the right. The first strategy is that the instructor can improve the design and delivery of the course. For example, the instructor might

Figure 1: Impact of an Improvement in Teaching Effectiveness



assign graded problem sets on a regular basis, increase the number of scheduled office hours, or conduct review sessions before each examination. Of course, as McKenzie noted, this strategy would be costly to instructor because it would reduce the amount of time available for other activities. A second option would be to institute a more generous grading policy. To be specific, I will assume that the instructor adds a uniform 10 points to each raw examination score while maintaining the same letter grade cutoffs. For example, a raw score of 70 would become an 80 and the student's letter grade would improve from a C to a B. I will assume that an instructor pursuing this strategy does not change the level of effort that s/he devotes to the course. The instructor continues to teach the same topics at the same depth. In assigning raw scores to examinations, the instructor does not change his/her standards. The raw scores represent a good faith effort on the part of the instructor to assess the extent to which the students have mastered the subject matter of the course. When assigning numerical grades, the instructor gives students the same level of feedback as before. The instructor's third option is to reduce the rigor of the course. For example, the instructor can eliminate difficult topics from the course syllabus and switch from essay examinations to multiple choice or true/false examinations where the questions have obvious answers. In this case, the instructor would be devoting less time and effort to the course.

PLAN A: IMPROVE TEACHING EFFECTIVENESS

If the instructor improves his/her teaching effectiveness, the transformation curve for each type of student will rotate to the right as shown in figure 1. For any given input of time, a type A student will receive a higher grade and a type B student will learn more. Each type of student will reach a higher indifference curve. The increase in overall student welfare will result in an improvement in the instructor's teaching evaluations.

In figure 1, each type of student will consume more leisure at the new welfare maximizing point. This is not the only possible outcome. The rightward rotation of the transformation curve produces effects that are analogous to the income and substitution effects that accompany a price change. Because the opportunity cost of earning a higher grade (or learning more) is lower, the substitution effect will lead the student to consume less leisure time. Assuming that both leisure and grades (or learning) are normal, the income effect will lead the student to consume more leisure time. The overall change in the amount of leisure consumed is ambiguous and depends on which effect is stronger. [See Lichty et al. (1978) for a more detailed discussion of the income and substitution effects.]

The observed result is that the average grade in this course will increase along with the instructor's SET rating. To the extent that the instructor is "buying" better student evaluations, s/he is paying for them by working harder. Even though each type of student is consuming more leisure, each type of student is learning more. In essence, the students have substituted the instructor's increased effort for their own.

There is evidence in the literature that the best formula for improved student evaluations is for the instructor to improve the delivery of the course. A number of papers have explored the instructor qualities that students value. The results are reassuring. Boex (2000) reports that "from the student's point of view, organization and clarity was the single most important attribute of effective economics education (p. 213)." Bosshardt and Watts (2001) find that students care about enthusiasm and preparation. DeCanio's (1986) results highlight the importance of organization and structure.

In papers that include both teacher qualities and grades (or expected grades) as explanatory variables, the instructor qualities turn out to be the most important determinants of SET scores. For example, Boex found that a one point improvement in his organization and clarity measure lead to a one point increase in SET scores in undergraduate core and noncore courses. On the other hand, giving students better than expected grades only increased the SET score by .29 of a point. Bosshardt and Watts report a similar finding.

While inflating grades might be a less effective method of improving student evaluations, it is also a lower cost method. For a time-constrained faculty member, this could well be an important consideration.

PLAN B: INFLATE GRADES

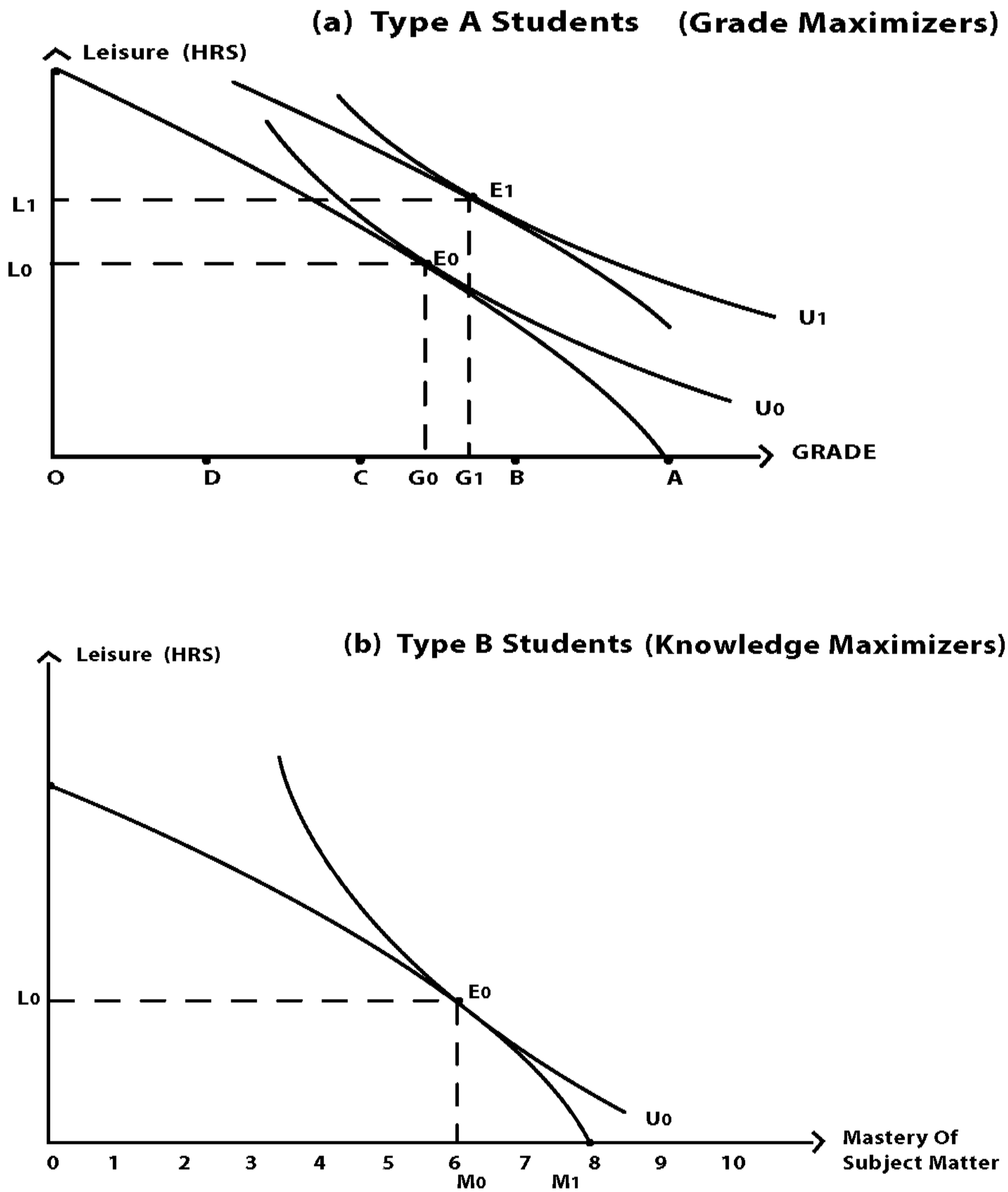
In this section, I am assuming that the instructor relaxes her/her grading standards but doesn't change the effort that s/he devotes to the course. In the case of grade-centered students, this policy change will shift the transformation curve to the right in a parallel manner as shown in figure 2. Previously, if a student devoted no time to the course s/he would earn an E, which corresponds to the origin. Now with no input of his/her own time the student can earn a D. Before, in order to earn an A, the student would have to devote all of his/her time to the course. Now, the student can earn an A and still enjoy some leisure time. As long as leisure and grades are both normal, a type A student will increase his/her consumption of leisure and earn a higher grade in the course. As Lichty et al. (1978) and McKenzie (1975) both observe, the observed grade for a type A student will increase by less than a grade level. Type A students are taking part of the benefits of grade inflation in the form of more leisure. Still, the increase in student welfare will lead to an improvement in the instructor's evaluations.

In the case of type B students, who only care about how much they learn, the transformation course will be unaffected since the change in grading standards does not make type B students more effective learners. Their welfare maximizing point won't change. In the absence of an increase in their welfare, type B students will not rate the instructor more highly.

The overall impact on the observed mean class grade and the instructor's SET rating will depend on the distribution of the class between type A and type B students. If the entire class consists of type A students, the mean course grade will increase by less than a grade level and the instructor will get better student evaluations. If the class is composed entirely of type B students, the mean course grade will increase by exactly a grade level, because these students are still devoting the same amount of time to the course. Thus, the less that students care about grades, the larger the increase in the class mean grade.

The improvement in the instructor's SET score depends on the division of the class between the two student types. As the mix of students shifts towards type B students, the increase in the instructor's SET rating will become smaller. The implication is that the results of an attempt to empirically estimate the impact of grades (or expected grades) on SET scores will depend on the setting in which the study takes place. A reasonable hypothesis is that grade inflation would produce the largest payoff in introductory, general education courses in which the majority of the students are taking the course to fulfill a distribution

Figure 2: Impact of a Reduction in Grading Standards



requirement rather than out of an interest in the subject matter. Since these students do care about their GPAs, they will value the improved grade and the additional leisure. In upper division classes, the share of type B students is likely to be larger. Upper division courses tend to be taken by majors who have an interest in the subject matter of the course. In addition, because they have taken several courses in the subject, they probably have a better ability to evaluate the teacher's performance. There are hints in the literature that this is the case. McPherson et al. (2009) find that if instructors of principles of economics courses can raise their students' expected grade by one grade point their SET score will improve by .27 of a point. In intermediate level courses, the same improvement in expected grade would only lead to a .1 of a point improvement in the SET score.

This analysis also has implications for the end users of SET ratings. Consider two instructors of principles of microeconomics: Dr. X attempts to improve his student evaluations by devoting more effort to teaching the course while Dr. Z resorts to a more generous grading formula. Since type A students are likely to constitute the majority of students, both instructors should see an improvement in the SET scores accompanied by an increase in the mean class grade. The students in Dr. X's course should learn more, while the overall level of learning in Dr. Y's course will decrease because students are devoting less time to the course and Dr. Y is devoting no additional time to the course. In the absence of an independent measure of student learning, there is no way to differentiate the two approaches on the basis of course grades and SET ratings. In this situation, a careful examination of the other material in the teaching portfolio, especially the course materials and self-statement, would be warranted.

In my department, we expect the faculty to explain deviations between their grades and those of the department and any marked changes in their grading pattern. Dr. X would be able to include material in his teaching portfolio such as copies of homework problem sets and the accompanying answer guides, examination review sheets, and class handouts that demonstrate that s/he is making a larger time commitment to the course. Dr. X would also be able to discuss these changes in his/her teaching statement. It is my experience that members of our Promotion and Tenure Committee tend to skim over course materials unless a candidate gives us a compelling reason to take a closer look at them. Dr. Z would have a more difficult time providing a convincing explanation for the increase in the mean grade in his/her courses.

An alternative to curving examination grades would be to include easy-to-earn bonus points, say for class participation, in the grading formula. The main drawback to this approach is that the grading formula must be included in the course syllabus. In this case, the participation points might catch the attention of some readers of the dossier who might view them as an overt attempt to improve their SET scores by inflating grades. The advantage of curving examination grades is that curving schemes are typically not disclosed in the course syllabus and cannot easily be detected by an outside reader of the dossier. An instructor who wishes to increase student grades without giving the appearance of attempting to "buy" better student evaluations could pursue a mixed strategy. The instructor could make some improvements in the delivery of his/her courses, such as conducting review sessions before examinations and scheduling more office hours, in addition to grading more leniently. In this case the course improvements would provide a justification for the increase in grades.

PLAN C: REDUCE THE RIGOR OF THE COURSE

In this case, the instructor attempts to improve his student evaluations by reducing the rigor of the course. This is the case where differences in student motivations make the most difference.

As shown in figure 3, the transformation curve for a type A student will shift to the right. The student will experience an increase in utility and be inclined to rate the instructor more highly. Because the student is devoting less time to the course and the course has been watered down, the student will learn less but won't care.

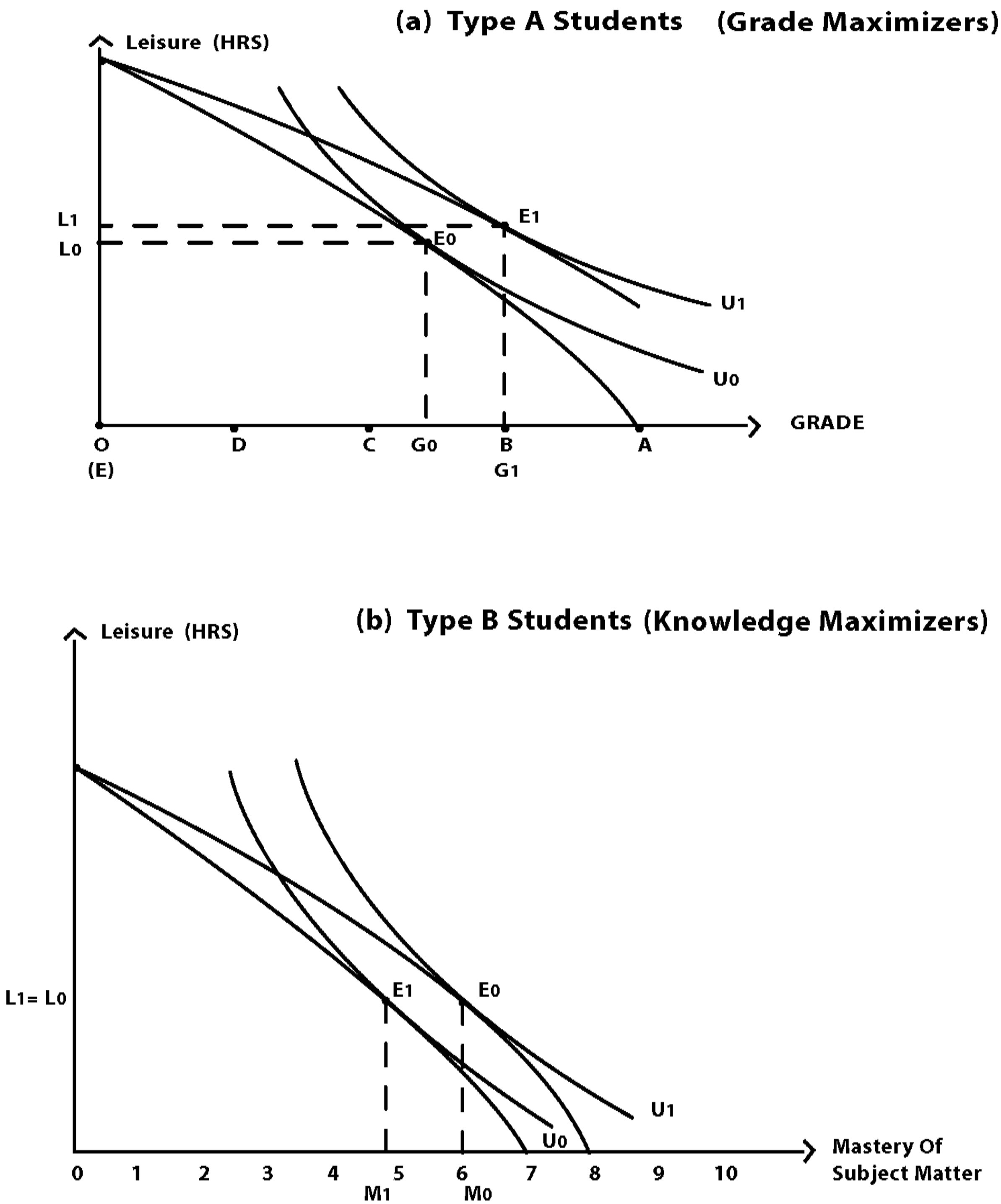
In the case of a type B student, the transformation curve will shift inward. For each level of leisure, the student will learn less than before, or to achieve a given level of learning the student will have to commit more time to the course. This would result in written comments such as: "Professor Y is so bad that I had to teach myself the course." The inward rotation of the transformation curve would place a type B student on a lower indifference curve. The reduction in the student's satisfaction would lead to a lower instructor rating.

Again, the impact on the instructor's SET rating depends on the mix of students in the course. The difference between Plan B and Plan C is that with Plan B the instructor's SET rating can only improve. In the case of Plan C, the SET rating can either increase, remain the same or even decrease depending on the division of the class between type A and type B students. Whether reducing the rigor of the course is a winning strategy for improving student evaluations is likely to depend on the level of the course. In an introductory course, an improvement in student evaluations is probable. In an advanced course such as econometrics or auditing, where type B students are likely to predominate, the instructor's reward would likely be less favorable evaluations since the students would feel that the course leaves them less well prepared for graduate school or the CPA examination.

An increase in rigor in the proper setting could lead to an improvement in an instructor's evaluations as long as the students are convinced that the course is helping prepare them for their post-graduate careers. Of course, an increase in rigor in a class dominated by type A students would lead to lower evaluations. At my college, students expect that general education classes should be easy. Instructors who attempt to enrich their general education courses by requiring their students to do more reading and writing can expect to read student comments that state the class is "too hard for an intro level course."

This result has implications for consumers of SET data. Consider a third instructor of principles of microeconomics, Dr. Y, who attempts to buy better SET scores by watering down his course. Since type A students are likely to constitute the majority of the students, Dr. Y's grade distribution data and SET scores should be roughly similar to those of Dr. X and Dr. Z. Evidence of the reduced rigor of Dr. Y's course would be found by a close reading of his/her examinations, which would be noticeably easier than those of Drs. X and Z. Of course, this close reading would place another demand on the time of the members of the Promotion and Tenure Committee or the members of department personal committees. Ideally, we would hope that peer evaluators would also comment on the lack of rigor.

Figure 3: Impact of a Reduction in Course Rigor



The distribution of students between the courses of Drs. X, Y and Z should not be viewed as exogenous. At most colleges, students have some ability to choose the course sections they sign up for. In making their selections, students have access to on-line sources of information such as “Rate My Professor” and the experiences of their classmates. Isely and Singh (2005) raise the possibility that an "instructor's reputation could attract high achievers or low performing students may be induced to drop the class (p. 29)." If this hypothesis is correct, then we would expect type B students to gravitate to Dr. X's sections. While the transformations curves have the same basic shape for both types of students, it seems reasonable to assume that the transformation curves for the typical type B student should lie to the right of the transformation curve for the typical type A student, The implication is that for a given time input the typical type B student would learn more and earn a higher grade. The result is that in Professor X's courses the larger representation of type B students should lead to better student performance and better course evaluations. Given the self-selection of the part of students, this positive correlation could not be interpreted as an attempt on the part of Dr. X to "buy" better evaluations. The possibility that student self-selection could lead to biased inferences has been explored in the literature, although as McPherson (2006) notes the results have been mixed. Here it should be noted that while students may well have preferences regarding the instructor with whom they would like to take a course, their ability to act on their preference depends on how early they get to register. Thus, a student who prefers Dr. Z might wind up in Dr. X's section because the only seats that are available are in Dr. X's section.

CONCLUSION

Allowing for differences in student motivation has important implications for both the empirical analysis of the determinants of SET scores and the use of SET data by the members of committees that evaluate faculty performance.

Introducing differences in student motivation into the analysis provides an additional explanation for the failure of empirical analyses of the determinants of SET ratings to find a consistent relation between SET scores and grades or expected grades. This paper identifies plausible conditions under which an increase in the average course grade (or expected course grade) could either lead to an increase, decrease or no change in SET scores depending on the strategy an instructor employ. The level of the course matters. A successful strategy to improve SET ratings in a principles course might not work in an upper level course.

The end users of SET data should not focus their attention solely on SET data or even on SET data in conjunction with data on grades. This data may not tell a clear story about what is actually going on in the classroom. The remaining elements of the teaching portfolio are intended to provide the evaluators of faculty performance with a more nuanced picture of an instructor's performance. They deserve at least as much attention as the SET data and grade distribution data.

An interesting follow-up to this paper would be to follow the lead of Grimes et al. (2004) and carry out a survey aimed at determining the extent to which students are motivated by a desire to master the subject matter of a course. The survey should be administered in a variety of courses from the introductory to the advanced levels. The survey should also address two other issues: the willingness of students to trade less learning for a better grade and whether they would be willing to reward the instructor with better evaluations. Conceivably, students might regard the time the instructor saves by watering down a course as a sufficient payoff and not see the need to further reward the instructor with improved SET scores.

REFERENCES

Becker, W.E., W. Bosshardt, and M. Watts. 2012. How departments of economics evaluate teaching. *Journal of Economic Education* 43(3): 325-333.

Becker, W.E., and J. R. Powers. 2001. Student performance, attrition, and class size given missing student data. *Economics of Education Review* 20(4): 377-388.

Bedard, K., and P. Kuhn. 2008. Where class size really matters: Class size and student ratings of Instructor Effectiveness. *Economics of Education Review* 27(3): 253-265.

Boex, L.F.J. 2000. Identifying the attributes of effective economics instruction: An analysis of student evaluations. *Journal of Economic Education* 31(3): 211-227.

Bosshardt, W., and M. Watts. 2001. Comparing student and instructor evaluation of teaching. *Journal of Economic Education* 32(1): 3-17.

Carrell, S.E., and J.E. West. 2010. Does professor quality matter? Evidence from random assignments of students to professors. *Journal of Political Economy* 118(3): 409-432.

DeCanio, S.J. 1986. Student evaluation of teaching: A multinomial logit approach. *Journal of Economic Education* 17(3): 165-176.

Dilts, D.A. 1980. A statistical interpretation of student evaluation feedback. *Journal of Economic Education* 11(2): 10-15.

Finegan, T.A., and J.J. Siegfried. 2000. Are student ratings of teaching effectiveness influenced by instructor's English language proficiency? *The American Economist* 44(2): 17-29.

Grimes, P.A., M.J. Millea, and T.W. Woodruff. 2004. Grades-who's to blame? Student evaluation of teaching and locus of control. *Journal of Economic Education* 35(2): 129-147.

Hammermesh, D.S., and A. Parker. 2005. Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review* 24(4): 369-376.

Isely, P., and H. Singh. 2005. Do higher grades lead to favorable student evaluations? *Journal of Economic Education* 36(1): 29-42.

Kanagaretnam, K., R. Mathieu, and A. Thevaranjan. 2003. An economic analysis of the use of student evaluations: Implications for universities. *Managerial and Decision Economics* 24(1): 1-13.

Kelley, A.C. 1972. Uses and abuses of course evaluations as measures of educational output. *Journal of Economic Education* 4(1): 13-18.

Kjelland, J. 2008. Economic returns to higher education: Signaling v. human capital theory an analysis of competing theories. *The Park Place Economist* 16: 70-77.

Lichty, R.W., D.A. Vose, and J.M. Peterson. 1978. The economic effects of grade inflation on instructor evaluation: An alternative interpretation. *Journal of Economic Education* 10(1): 3-11.

Love, D.A., and M.J. Kotchen. 2010. Grades, course evaluations, and academic incentives. *Eastern*

Economic Journal 36(2):151-163.

McKenzie, R.B. 1975. The economic effects of grade inflation on instructor evaluations: A theoretical approach. *Journal of Economic Education* 6(2): 99-105.

McPherson, M.A. 2006. Determinants of how students evaluate teachers. *Journal of Economic Education* 37(1): 3-20.

McPherson, M.A., R.T. Jewel, and M. Kim. 2009. What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal* 35(1): 37-51.

Needham, D. 1978. Student effort, learning and course evaluations. *Journal of Economic Education* 10(1): 35-43.

Nelson, J.P., and K.A. Lynch. 1984. Grade inflation, real income, simultaneity, and teaching evaluations. *Journal of Economic Education* 15(1): 21-37.

Ragan, J.F., and B. Walia. 2010. Differences in student evaluations of principles and other economics courses and the allocation of faculty across courses. *Journal of Economic Education* 41(4): 335-352

Weinberg, B.A., M. Hashimoto, and B.M. Fleisher. 2009. Evaluating teaching in higher education. *Research in Economic Education* 40(3): 227-261.

Weiss. A. 1995. Human capital vs. signaling explanations of wages. *The Journal of Economic Perspectives* 9(4): 133-154.

Zangenzadeh, H. 1988. Grade inflation: A way out. *Journal of Economic Education* 19(3): 217-230.

NBA Coaching Changes: The Role of Market Expectations, Race, and Former Players

Rodney J. Paul*, Nick Riccardi†, Andrew Weinbach‡, Mark Wilson§

ABSTRACT

Coaches in sports are similar to Chief Executive Officers. Their retention or dismissal is not only influenced by actual performance, but performance compared to expectations. In the National Basketball Association, like other professional sports, a futures market exists that allows for market-based expectations of team performance in the form of win totals. We utilize these win totals to compute performance compared to expectations and use it in logit models of coach retention and dismissal. While performance compared to expectations is found to play a statistically significant role in retention and dismissal, race and former player status is not statistically significant.

I. INTRODUCTION

The head coach of a sports team is a highly visible and important figure in the sports world. Their expected tenure in the position is not long as turnover occurs frequently across different sports leagues and organizations. Teams often dismiss their coach to attempt to improve team performance, as it is easier to replace the person organizing and running the team, than the players themselves.

In many ways, a head coach shares similar responsibilities with a corporate Chief Executive Officer (CEO). Coaches are hired to win games with the roster they are given by their general managers and ownership. Coaches are responsible for the overall performance of their organization and often pay the price for suboptimal performance. Dismissal most frequently occurs when a team does not have a good season, specifically we assume this is when they do not meet expectations of ownership. Dismissals of coaches could occur due to other factors, such as retirement, illness, bad behavior, or violation of contract. For the purposes of this paper, we assume that any dismissal of coaches which occurs during or after the season is due to performance issues of the team, except where otherwise noted. In the professional sports world, coaches do not typically voluntarily leave one team to coach another, which typically does happen in college sports. Therefore, any time a coach is dismissed by management, their contract not renewed, or is released from their contract, we assume it is performance-based.

This study investigates the dismissal or retention of coaches in the National Basketball Association (NBA). There is not a specific organizational structure for every NBA team, as roles within an ownership group, front office, and coaching staff vary from team-to-team. The main decision-making personnel are the owner or ownership group, the general manager, and the head coach. Owners can range from taking a

* Syracuse University, NY, Email: Rpaul01@syr.edu
† Syracuse University, NY
‡ Coastal Carolina University, SC
§ St. Bonaventure University, NY

hands-off approach, leaving decisions regarding roster creation, draft selection, and coaches to their general managers to having the complete opposite approach, being very hands-on, in relation to anything from making picks on draft day to negotiating trades with other teams.

General managers are thought to have the largest role in roster creation as it relates to signing players and making trades. Typically, general managers also decide on the head coach and coaching staff either independently or in conjunction with ownership. Head coaches might have input into how a roster is constructed but are responsible for earning wins based on the talent made available to them by the front office.

The specific title of general manager on an NBA team varies across teams. Some are directly titled “General Manager” while others are referred to as “President of Basketball Operations.” Two examples of the structure of an NBA front office can be seen from the current (2023) iterations of the Boston Celtics and Los Angeles Lakers. The Celtics have an ownership group comprised of Wyc Grousbeck, H. Irving Grousbeck, Stephen Pagliuca, and the Abbey Group. The team president is Rich Gotham, Brad Stevens is the President of Basketball Operations (the equivalent of General Manager on other teams), and Joe Mazzula is the head coach. The Lakers are owned by Jeanie Buss, who simultaneously serves as President of the organization as well. Rob Pelinka is the General Manager and Vice President of Basketball Operations, while Darvin Ham serves as Head Coach.

We assume the role of the coach is to provide winning seasons to the organization with the roster given to them by the front office. As it relates to performance itself, coaches are again similar to CEOs as the absolute level of success is often less important than the relative level of success achieved by coaches. Two coaches who have the same win percentage may be viewed quite differently by their respective front offices. This is due to the importance of expectations of performance. If expectations are low, coaches achieving even mediocre success may be rewarded for their actions. On the other hand, winning more often than losing is not enough when expectations are high. Coaches, like CEOs, are often dismissed even if their level of success is beyond most of their peers, specifically, when they disappoint compared to market expectations.

To thoroughly investigate coaching performance in the NBA, a proxy for expectations is needed. While stock market forecasts serve as a good proxy for expected CEO performance, a similar market exists in the sports world in the form of the betting market. Specifically, there are preseason markets for win totals for the upcoming season posted at sportsbooks domestically and around the world. These market-based prices provide expectations of team performance in the upcoming season. An NBA season consists of eighty-two games. Win totals are a proposition bet where participants in the marketplace wager money that a team wins more or fewer games than the posted number. When all available information is incorporated into this price, it should serve as a good forecast of future events. Therefore, these prices serve as a measure of expectations and allow for a simple comparison of actual performance compared to expected performance to help to explain when and why coaches are retained in their position or replaced by a different coach.

In addition to examining the role of expectations in this marketplace, studying the retention and dismissal of coaches allows for an investigation of biases. A key element of both academic research and public interest is coaches and race. Past studies of the National Football League (NFL) showed bias against Black coaches in the sport. The NFL and other sports have policies in place to specifically try to increase hirings of minority coaches. Our data set allows for the investigation of racial bias as we directly evaluate if, all else equal, minority coaches are dismissed at a higher rate than their white counterparts.

Another potential bias we investigate in this study relates to bias toward former NBA players as coaches. As in other sports, former players become coaches at various levels of play. Some become NBA coaches, overseeing an organization where they recently were one of the players. These former players may have advantages over other coaches in the league due to their popularity from their playing days. We directly assess if former NBA players are less likely to be dismissed, all else equal, compared to their peers due to their presumed popularity with the local fan base.

We evaluate for the role of performance, performance compared to expectations, and potential racial and former player biases through logit models of coaching performance. We interpret the results and investigate potential rationales for the findings. The paper is structured as follows. Section II contains a brief literature review on CEO turnover and sports-related coaching performance studies. The third section presents the data and role of the betting market in expectations. The fourth section presents the logit model results as it relates to coach dismissal, expectations, race, and former player status. The concluding section discusses the results and offers conclusions to the study.

II. LITERATURE REVIEW

Past literature explores the link between head coaches of sports teams and corporate CEOs. Theoretical foundations for Chief Executive Officer turnover were explored in studies by Frederickson, et al. (1988), Franck, et. al. (2010), Holmes (2010), and Frick, et al. (2010). Frederickson, et al. (1988) develops a model of expectations and performance. They focus on the relative power of the incumbent CEO, the individuals in charge of retention and firing, and the availability of viable alternatives in their model, in addition to the actual performance compared to market expectations. Franck, et al. (2010) spotlights the risk-taking aspects of turnover in CEOs in their study. A Bayesian learning model of estimating true abilities of CEO and a cost-benefit approach to retention and firing is utilized by Holmes (2010). Frick, et al. (2010) adds that CEO turnover is consistent with standard principal-agent theory. A common thread in all of the studies is performance compared to expectations plays a key role in the decision to fire or retain a CEO.

A common method in the literature to estimate expected performance for CEOs and managers is through use the of industry analyst forecasts. Farrell and Whidbee (2003) discuss the advantages and disadvantages of this approach in their study. In the corporate setting, performance expectations based on industry analyst forecasts are a significant determinant of CEO turnover (Brickley, 2003).

In sports, the head coach or manager of a team is viewed as analogous to corporate CEOs, as their retention or dismissal also depends on performance compared to expectations. Wangrow, et al. (2018) use survival analysis to examine dismissal of NBA coaches. Their model includes current performance and measures of expectations including relative salaries and previous year's performance. They find that coaches that underperform expectations are more likely to be dismissed but find no evidence that minority coaches are treated differently. In their analysis of college football, Humphreys, et al. (2016) use a logit model indicating whether a coach is dismissed at the end of the season and find that coaches with higher cumulative winning percentages against the point spread, a measure of performance against expectations, are less likely to be dismissed. Salaga and Juravich (2020) study the length of NFL head coach employment from 1985-2018 as measured by the total number of regular season games the individual retains employment at the position. Using a Cox hazard model, they find that a greater winning percentage decreased the probability of being dismissed, and winning percentage against the point spread significantly impacted the probability of dismissal for low performing head coaches, though not for high performing head coaches.

Three studies, Buraimo, et al. (2017), Elaad, et al. (2018), and Pieper, Nuesch, et al. (2014) incorporate betting odds to evaluate coaching performance in European soccer. These studies uncover performance compared to expectations play a significant role in the likelihood of coach retention or dismissal. Coaches that underperform betting market expectations for points are more likely to face dismissal. In Dutch professional soccer, van Ours and van Tuijl (2016) examine in-season coaching changes and also find that underperforming betting market expectations increases the likelihood of head coach dismissal, but a coaching change does not improve team performance. Barros, et al. (2009) analyze head coaching dismissals in the German Bundesliga using a hazard model including information on team wage bills and head coach salaries as an indicator for fan and management expectations and find that higher payrolls lead to earlier dismissals.

Using data from Sportsoddshistory.com, Roach (2020) reveals that for Major League Baseball, teams with higher win totals and teams that underperform expectations tend to increase spending on payroll to a greater degree. Gomez-Gonzalez et al. (2018) use the probabilities of winning games from betting odds to calculate the efficiency (actual wins vs. expected wins over the course of a season) of NBA coaches and find that Black NBA coaches are more likely to be fired for similar levels of efficiency than white NBA coaches. Del Corral, Maroto, & Gallardo (2017) in their analysis of coaching efficiency in the Spanish Basketball Club Association (ACB) find evidence that being a former player had a positive and significant impact on efficiency in one of three models in their Stochastic Frontier Cobb-Douglas production function analysis, though OLS estimates of betting odds efficiency determinants, the ex-professional player dummy variable is not found to be statistically significant. Holmes (2010) uses a hazard model to examine head coaching dismissals in college football and finds that stronger recent performance decreases the chance of dismissal but stronger past performance increases expectations of future performance, which increases chance of dismissal.

The role of discrimination is explored in a variety of settings and studies in the literature, Heilman and Caleo (2018) distinguish between access and treatment discrimination. Access discrimination occurs ex-ante in the process of hiring. Certain groups are faced with limitations to their hiring that is not based on actual or expected performance in the role. This is typically applied to race and gender discrimination. On the other hand, treatment discrimination is ex-post, occurring after a hire has been made. It postulates that certain groups are not treated fairly, such as being dismissed sooner, based on specific characteristics such as race and/or gender. Cunningham (2019) expands these concepts as it relates to diversity in the workplace.

Madden (2004) examines discrimination against minority NFL head coaches using a logit model and finds that minority coaches were dismissed earlier than white coaches with similar performance characteristics. A later study by Madden and Ruther (2011) using similar methodology does not find evidence of discrimination. The authors note that the introduction of the Rooney Rule, a policy that requires NFL teams to interview ethnic-minority candidates for head coaching and senior football operations positions, is the reason that discrimination against minority coaches is not statistically significant in their sample. Salaga and Juravich (2020) do not find discrimination against NFL minority coaches using a model that includes a measure of expectations based on individual game results of the head coach against-the-spread along with other factors.

III. COACHING DATA AND THE ROLE OF THE BETTING MARKET IN EXPECTATIONS

The data for this study on team performance and head coaches is compiled from the website www.basketball-reference.com, which has season-by-season performance data for the NBA. For each season, it lists the head coach, or coaches, of each team in the league. Personal bios on the coaches available on the site are used to note if the coach was a former NBA player. Race of NBA coaches is taken from the Wikipedia entry on the history of minority coaches in the NBA which is found at: https://en.wikipedia.org/wiki/Category:African-American_basketball_coaches.

A key element of this study is the computation of performance of coaches compared to expectations. Actual win percentages are directly compiled from the basketball-reference.com season page. Win totals are obtained for each team and season in the sample from www.sportsoddshistory.com. The season total is an over/under bet on the number of games an NBA team wins in a season. Season totals are set in advance of the season and fluctuate due to added information, bettor actions, and sportsbook manager actions during the off-season. We use the season total from before the start of the regular season as compiled by sportsoddshistory.com.

Our sample consists of every NBA coach who started a season from the 1999-00 season through the 2018-19 season. The data for during the pandemic years are available on the website, but is not used due to the delays, shortening of season, and unique circumstances in which NBA games were played during this time. Coaches dismissed mid-season are included in the sample, using the actual win percentage of the team at the time of firing compared to the expected win percentage from the season totals betting market. Replacement coaches hired during the season are not included in the sample as we do not have updated futures for each new in-season hire at the time they assumed the role of coach of the team.

Our assumption about the wagering market season win totals is that they serve as a market forecast of team performance for the upcoming season. How teams and coaches actually perform against these expectations are measured and included in the model to account for how good or poor a team is expected to perform. We assume management hires a coach with the goal of maximizing wins, reaching the playoffs, and winning championships. Therefore, the coach strives to get the most out of the roster of players to win games and put the team in a position to earn a playoff berth and have the opportunity at a championship. Whether a coach is retained or dismissed is dependent not only on their actual performance, but how they fared compared to market expectations. This is due to management expecting coaches to at least meet expectations and we stipulate the betting market wins total serves as an excellent proxy for management expectations as it relates to season performance.

For example, an NBA team that won forty-nine games in the regular season may or may not be looked up on favorably, even though they won more often than they lost in an 82-game season. If the team was expected to be a .500 team (41 wins), the team easily exceeded expectations and the coach's job is typically safe. On the other hand, if the team was expected to win fifty-five games, but they only won forty-nine, then the team and coach disappointed, with the coach having an increased probability of being dismissed and replaced. Therefore, the betting market season win totals are likely to provide information which is of considerable value when it relates to understanding the motivation of front offices and their decision-making in the NBA. NBA owners and general managers are more likely to dismiss a coach when the team does not meet expectations and are more likely to retain a coach if the team and coach exceed expectations.

IV. COACH DISMISSALS, EXPECTATIONS, RACE, AND FORMER PLAYER STATUS

To test for the impact of performance compared to expectations and for any potential bias toward or against minority coaches and former NBA players as coaches, we construct a logit model. The dependent variable in the logit model is a dummy variable that takes the value of one if the coach is dismissed and a value of zero if retained (continues to coach with the team the following season). Reiterating from the introduction, we assume all dismissals are performance-related, so any coach who is outright fired, does not have contract renewed, or is released from their contract is considered dismissed as it relates to the dependent variable. All coaches who start an NBA season are included in the sample.

The independent variables in the logit model include win percentage, performance compared to expectations, tenure of coach with team, dummy variables for minority coach and former NBA player as coach, and season dummies. The win percentage is the simple win/loss percentage of a coach for the season. If the coach was dismissed or replaced before completion of the season, the win percentage is calculated at the time the replacement was made. Expectations on the coefficient on win percentage is negative as better team performance leads to a reduced chance of being dismissed.

Our proxy for performance compared to expectations is the actual win percentage of the team minus the expected win percentage of the team. The expected win percentage is calculated from the season win total in the betting market which was obtained from the website sportsodshistory.com. This variable is expected to have a negative effect on the chances of being dismissed or replaced. If a coach outperforms expectations, the variable takes a positive value, and this should lead to a lower probability of being dismissed. If a coach disappoints compared to expectations, the variable takes a negative value, which should lead to a greater possibility of being dismissed and replaced as coach.

Games with team and its square are included in the model to account for coaching tenure with the team. New hires are likely to have more leeway in terms of success as they implement their strategies and personnel over time. After time with the team, however, their message and approach may grow stale with the team and its players, leading to their dismissal. Therefore, we expect a nonlinear relationship with this variable as it relates to the binary dependent variable of coach being dismissed or retained. Age of the coach at the start of the season is also included as an independent variable in the model. The birthday of each coach is compiled from www.basketballreference.com. Age is computed using the start date of each NBA season compared to the birth date of the coach.

To test for racial bias and bias toward or against former NBA players as coaches, simple dummy variables are constructed and included in the model. The list of minority head coaches in the NBA is taken from Wikipedia and any coach who is noted as a minority received a value of one for this variable, while all others were zero. In a similar fashion, former NBA players who were head coaches during our sample were identified from basketball reference (www.basketballreference.com) and their observations took a value of one for this variable, while coaches who were not former NBA players took a value of zero. Season dummy variables are also included in the model but are not found to offer much in the way of statistical significance individually or jointly. Full results of the individual season dummies are not shown in the table, to preserve space, but are available from the authors upon request.

Summary statistics for the non-binary variables are shown in Table 1 below.

Table 1: Summary Statistics of Non-Binary Variable

Variable	**Mean**	**Median**	**Standard Deviation**
Season Win Percentage	0.496	0.512	0.158
Expected Win Percentage (Based on Season Win Total)	0.505	0.518	0.125
Season Actual Win Percentage Minus Expected Win Percentage	-0.010	-0.006	0.109
Number of Games Coached with Team	293.857	204.000	289.281
Coach's Age in Years	51.187	51.080	7.959

Before delving into the models, it is useful to explore the nature of tenure for Head Coaches in the NBA. Tenure is recorded in our sample as the number of consecutive seasons a coach started the season with their team. Summary Stats for coaching tenure are shown in Table 2.

Table 2: Coaching Tenure – Overall Sample – 1999-00 through 2018-19 Seasons

Minimum	1st Quartile	Median	Mean	3rd Quartile	Max
1.000	2.000	3.000	3.418	4.000	23.000

As seen in the table, the median coaching tenure is not long, three years, with the mean being slightly higher (3.4 years). The maximum tenure in the sample was twenty-three years, while the minimum was one.

Table 3 shows the breakdown of coaches by race and former NBA player status. The second column shows the number of coaching stints, which includes observations for the same coach with different teams, while the third column shows just the number of coaches in each category (does not count multiple stints as coach). Average coaching tenure by race and former NBA player status is included in Table 4.

Table 3: Number of Coaches by Race and Former Player Status in Sample

Category	**Coaching Stints** (Includes same coach for different teams)	**Coaches** (Coaches who have coached for one or more teams)
White	119	70
Black	73	43
Asian	1	1
Hispanic	1	1
Former Player	125	73
Non-Former Player	69	42

Table 4: Average Coaching Tenure by Group in Sample

Group	Average Coaching Tenure
White	3.714
Minority	2.947
Former Player	3.456
Non-Former-Player	3.348
White	3.714
Black	2.863
Asian	11.000
Hispanic	1.000
White Former Player	3.882
Minority Former Player	2.947
White Non-Former Player	3.490
Minority Non-Former Player	2.944

In this study, we assume that each season for each coach is an independent event as it relates to if the coach is retained or dismissed. While coaches coach with the same team over multiple seasons, we assume that each season is an individual event and coaches can be dismissed at any time, despite past success, due to the pressures to win at the professional level of sports. With this assumption, we use a logit model where the dependent variable is if the coach is retained (observation of 0) or dismissed (observation of 1). Correlation coefficients and their respective p-values are shown in Appendix 1. Given concern over the observations not being independent, we also ran a Generalized Estimating Equation (GEE) model with clusters by coach using various correlation structures as an alternative model, which all yielded comparable results in terms of statistical significance of the independent variables using the computed robust z-statistics. The GEE with cluster by coach and independent correlation structure is presented in Appendix 2.

Three logit model specifications are shown in Table 5 below. The first model does not include either the minority coach or former NBA player dummy variables. Model II includes the minority coach variable, while model III includes both the minority and former NBA coach dummy variables.

Table 5: Logit Model Results of NBA Coach Firing - Dependent Variable: Dummy for Coach Dismissed/Replaced

Variable	**I**	**II**	**III**
Intercept	-1.457	-1.636	-1.636
	(-1.497)	(-1.621)	(-1.621)
Season Win Percentage	-2.4606***	-2.427**	-2.429**
	(-2.594)	(-2.554)	(-2.554)
Season Actual Win Percentage minus Expected Win Percentage	-7.671***	-7.682***	-7.676***
	(-5.284)	(-5.286)	(-5.276)
Number of Games Coached with Team	0.002*	0.002*	0.002*
	(1.701)	(1.739)	(1.739)
(Number of Games Coached with Team)^2	-2.175e-06*	-2.19e-06*	-2.19e-06*
	(-1.911)	(-1.929)	(-1.928)
Coach's Age in Years	0.030**	0.032**	0.032**
	(1.989)	(2.089)	(2.048)
Minority Coach Dummy Variable		0.160	0.156
		(0.679)	(0.644)
Former NBA Player Dummy Variable			0.017
			(0.071)
Season Dummies	YES	YES	YES
McFadden R^2	0.2034	0.2036	0.2038
Number of Observations	595	595	595
Number of Dismissals	165	165	165

Coefficients for the variables in the logit model are shown with accompanying z-statistics. Statistical significance is noted with * notation, where *** indicates statistical significance at the 1% level, ** indicates statistical significance at the 5% level, and * indicates statistical significance at the 10% level.

Across all three model specifications comparable results are found for win percentage, performance compared to expectations, and tenure with team. Actual win percentage of the team in the most recently completed season was negative and statistically significant at the 1% level. The better a team performs, the less likely the coach is dismissed. Tenure with the team is found to have a nonlinear effect as the games with team variable was found to have a positive effect, while its square is negative. The key variable related to performance and expectations, the actual win percentage minus expected win percentage (from season totals in betting market), is statistically significant at the 1% level and has a negative effect on if the coach was dismissed. Outperforming expectations leads to a coach being retained, while disappointing compared to expectations leads to a greater chance of being dismissed as coach.

The dummy variables for minority coach and former NBA player are found to be positive, but statistically insignificant. While the coefficient suggests a possible slight bias against minority coaches and former players, these variables are not statistically significant. This reveals that actual performance, both in absolute terms and relative to expectations, appears to be driving the dismissal or retention of NBA coaches, not race or former player status. Average Marginal Effects from the logit models are shown in Table 6 below. Average marginal effects describe the average change in explanatory variables on the change in the probability of outcomes in logistic regression. Specifically, the coefficients represent the average change in probability when the variable x increases by one unit. From the results, on average, a coach who has a winning percentage 0.1 below expectations is around 11% more likely to be dismissed.

Table 6: Average Marginal Effects from Logit Model of NBA Coach Firing

Variable	I	II	III
Season Win Percentage	-0.363***	-0.357***	-0.358***
Season Actual Win Percentage minus Expected Win Percentage	-1.132***	-1.133***	-1.132***
Number of Games Coached with Team	1.625e-04	1.684e-04	1.687e-04
Coach’s Age in Years	0.004*	0.005*	0.005*
Minority Coach Dummy Variable		0.024	0.023
Former NBA Player Dummy Variable			0.003
Season Dummies	YES	YES	YES

Alternative model specifications are available in the appendices. Appendix 3 includes two additional explanatory variables, the metro population of the city (a measure of market size) and a dummy variable if team ownership changed hands that season. The metro population is included to see if front offices in larger markets approach coach retention and dismissal in a separate way than smaller markets; i.e. larger market team ownership and front offices may be less patient with coaches due to media attention. The dummy variable for ownership change is to test for the possibility that new ownership will immediately want their own coach, thereby dismissing the previous coach regardless of team performance during the previous year. The last column of results in Appendix 3 removes the years from the sample where ownership did change hands, reducing the sample by twenty-three team-seasons and reducing the number of dismissals included by five compared to the whole sample, and shows the results.

In all included alternative specifications, which include different combinations of these variables in the full and restricted samples, the additional variables of interest, metro population and dummy for new ownership, were not found to be statistically significant. The results related to the other variables in the models did not change in a substantive way, as the coefficients, their signs, and the levels of statistical significance did not change much at all. Overall, the size of the metro population and inclusion of a dummy for new ownership did not meaningfully change the results. Removal of the team-seasons where ownership changed similarly did not meaningfully change the results of the models shown above.

To further investigate performance of coaches by race and former player status, we use t-tests to compare separate groups. In each case, we note the win percentage, expected win percentage (from season win totals futures), and actual minus expected win percentage averages. First, we observe the differences between white and minority coaches, then former players and non-players, and the last group combines the two (white former players compared to minority former players). The Welch two-sample t-statistic is presented with the corresponding p-value. Results are shown in Table 7 below.

Table 7: Welch Two-Sample T-Statistics Comparing Race and Former Player Status

	White Coaches	**Minority Coaches**	**t-statistic**	**p-value**
Observations	385	210		
Average Actual Win Percentage	0.514	0.462	3.820	0.001
Average Expected Win Percentage	0.517	0.485	3.021	0.003
Average Actual minus Expected Win Percentage	-0.003	-0.022	2.019	0.044
	Former Players	Non-Players	t-statistic	p-value
Observations	385	210		
Average Actual Win Percentage	0.491	0.504	-0.960	0.338

Average Expected Win Percentage	0.504	0.507	-0.287	0.774
Average Actual minus Expected Win Percentage	-0.013	-0.003	-1.096	0.274
	White Former Players	Minority Former Players	t-statistic	p-value
Observations	226	159		
Average Actual Win Percentage	0.519	0.451	4.247	0.001
Average Expected Win Percentage	0.522	0.479	3.584	0.001
Average Actual minus Expected Win Percentage	-0.003	-0.027	2.034	0.043

The results of Table 7 illustrate considerable differences across coaching groups. In the sample period studied, white coaches had 5% higher win percentages than minority coaches, statistically significant at the 1% level. This is due to minority coaches beginning their tenure on teams with poor rosters compared to the rest of the league. The expected win percentage of teams coached by minorities was lower than that of teams with white coaches, reflecting this outcome.

Both minority and white NBA coaches underperform compared to expectations, which is due to the preference of bettors to wager on the over compared to the under. Minority coaches are found to underperform compared to expectations by 2.2%, significant at the 5% level. In general, across the sample, minority coaches are given opportunities with less talented rosters (reflected by the lower expectations in the betting marketplace), but disappointed overall compared to expectations on average in the sample.

In the second group, no statistical differences are found between former NBA players as coaches when compared to non-former NBA players. In the final grouping in Table 7, white former NBA player coaches had win percentages 6.5% higher than minority former NBA player coaches, which was statistically significant at the 1% level. White former NBA player coaches had significantly higher expectations in the sample, while minority former NBA player coaches are found to underperform expectations by 2.7%, statistically significant at the 5% level. Similar to the findings above, minority coaches tended to start their tenure with teams with lower expectations overall, which reflects a less-talented roster compared to other teams.

V. DISCUSSION AND CONCLUSIONS

NBA coaches have positions that are similar to CEOs of companies as they are judged based upon the overall success of their organization. Their success, however, should be measured in relative terms to properly account for expectations. Coaches or CEOs that exceed expectations are rewarded and are typically retained moving forward. Coaches or CEOs that disappoint compared to expectations, however, are often dismissed even if their absolute performance would be considered successful in other teams or organizations.

While companies have expectations which are surmised from industry analyst forecasts, the sports world has gambling market data which relates to season-long performance. Season win totals, established before the season begins and based in an open marketplace, provide expectations of how many games each team is expected to win. Using this information, coupled with actual performance and other variables, we study the market for retention and firing of NBA coaches using a logit model.

While actual overall performance is found to have a negative and statistically significant effect on whether a coach is dismissed, as expected, the role of expectations was also extremely important. Actual win percentage minus expected win percentage, taken from the projected season win totals in the betting market, has a negative and statistically significant effect, at the 1% level, on whether a coach would be dismissed. This illustrates that performance compared to expectations is important when evaluating front office decisions as it relates to coaches. When performance exceeds expectations, the coach is likely to be retained. When the team disappoints compared to expectations, however, the odds increase that a coach is dismissed.

In addition to investigating the role of expectations in the retention and dismissal of coaches in the NBA, we also used this study to address concerns of bias as it relates to racial discrimination and preferable treatment of former NBA players who became coaches. Dummy variables for minority coaches and for former NBA players as coaches are both statistically insignificant. No statistical evidence is found in terms of discrimination against minority coaches or toward former players. Treatment discrimination is not found in the NBA in this sample. As in other sports, intense competition drives decisions based upon merit, rather than through discriminatory actions, at least after a coach has been hired. Due to the nature of the data, we use, we cannot make any conclusions about access discrimination; biases that occur before coaches are hired. However, we do find an absence of treatment discrimination among coaches hired, and that expectations compared to actual performance plays a key role in front office decision making in the NBA.

REFERENCES

African-American basketball coaches. (2023, June 23). In *Wikipedia*. https://en.wikipedia.org/wiki/Category:African-American_basketball_coaches

Barros, C., B. Frick, and J. Passos. Coaching for Survival: The hazards of head coach careers in the German Bundesliga. *Applied Economics*, 41, 2009, 3303–11.

Brickley, J. A. (2003). Empirical research on CEO turnover and firm-performance: a discussion. *Journal of Accounting and Economics*, 36(1-3), 227-233.

Buraimo, B., Bryson, A., and Simmons, R. (2017). Time to go? Head coach quits and dismissals in professional football. Working Paper No. 10600, *Institute for the Study of Labor (IZA)*.

Cunningham, G. (2019). *Diversity and Inclusion in Sport Organizations: A Multilevel* Perspective (4th ed.). Routledge.

del Corral, J., Maroto, A., and Gallardo, A. (2017). Are former professional athletes and native better coaches? Evidence from Spanish basketball. *Journal of Sports Economics*, 18(7), 698-719.

Elaad, G., Jelnov, A., and Kantor, J. (2018). You do not have to succeed, just do not fail: When do soccer coaches get dismissed? *Managerial and Decision Economics, 39*(3), 269-274.

Farrell, K. A. and D. A. Whidbee (2003). Impact of firm performance expectations on CEO turnover and replacement decision, *Journal of Accounting and Economics*, 36(1-3), 165-196.

Franck, E., Nüesch, S. and J. Pieper (2010). How expectations affect managerial change. Working Papers 0139, University of Zurich, *Institute for Strategy and Business Economics (ISU)*.

Frederickson, J. W., D.C. Hambrick, and S. Baumrin, (1988). A model of CEO dismissal. *Academy of Management Review*, 13, 255-270.

Frick, B., Barros, C., and Prinz, J. (2010). Analyzing head coach dismissals in the German “Bundesliga” with a mixed logit approach. *European Journal of Operational Research*, 200(1), 151-159.

Gomez-Gonzalez, C., del Corral, J., Maroto, A., and Simmons, R. (2018). Racial differences in the labor market: The case of expected performance and dismissals of head coaches in NBA. Working Paper 2018-2. DAEF, UCLM.

Heilman, M. and Caleo, S. (2018). Combatting Gender Discrimination: A Lack of Fit Framework. *Group Processes & Intergroup Relations*, 21(5), 725-744.

Holmes, P. (2010). Win or go home: Why college football coaches get dismissed, *Journal of Sports Economics*, 11(1), 1-22.

Humphreys, B., Paul, R., and Weinbach, A. (2016). Performance expectations and the tenure of head coaches: evidence from NCAA football. *Research in Economics*, 70(3), 482-492.

Madden, J. (2004) Differences in the success of NFL coaches by race, 1990-2002: Evidence of last hire, first fire. *Journal of Sport Economics*, 5(1), 6-19.

Madden, J. and Ruther, M. (2011). Has the NFL’s Rooney Rule efforts “leveled the field” for African American head coach candidates? *Journal of Sport Economics*, 12(2), 127-142.

Pieper, J., Nüesch, S., and Franck, E. (2014). How performance expectations affect managerial change. *Schmalenbach Business Review,* 66(1), 5-23.

Roach, M. A. (2020). Team performance relative to expectations and its impact on player personnel decisions in Major League Baseball. Available at SSRN 3701851.

Salaga, S. and Juravich, M. (2020). National Football League head coach race, performance, retention, and dismissal. *Sport Management Review*, 23(5), 978-991.

van Ours, J. C., and van Tuijl, M. A. (2016). In-season head-coach dismissals and the performance of professional football teams. Economic Inquiry, 54(1), 591-604.

Wangrow, D. B., Schepker, D. J., and Barker III, V. L. (2018). Power, performance, and expectations in the dismissal of NBA coaches: A survival analysis study. *Sport Management Review*, 21(4), 333-346.

APPENDIX

Appendix 1: Correlation Coefficients of Independent Variables

	Win Percentage	Expected Win Percentage	Actual minus Expected Win Percentage	Former Player Dummy Variable	White Dummy Variable	Minority Dummy Variable
Win Percentage	X	0.73 (0.000)	0.62 (0.000)	-0.04 (0.328)	0.16 (0.000)	-0.016 (0.000)
Expected Win Percentage	0.73 (0.000)	X	-0.09 (0.027)	-0.01 (0.767)	0.12 (0.003)	-0.012 (0.003)
Actual minus Expected Win Percentage	0.62 (0.000)	-0.09 (0.027)	X	-0.04 (0.281)	0.09 (0.036)	-0.09 (0.036)
Former Player Dummy Variable	-0.04 (0.328)	-0.01 (0.767)	-0.04 (0.281)	X	-0.17 (0.000)	0.17 (0.000)
White Dummy Variable	0.16 (0.000)	0.12 (0.003)	0.09 (0.036)	-0.17 (0.000)	X	-1.00 (0.000)
Minority Dummy Variable	-0.016 (0.000)	-0.012 (0.003)	-0.09 (0.036)	0.17 (0.000)	-1.00 (0.000)	X

Appendix 2: Alternative Model of GEE Estimates with Dismissed as Dependent Variable, Clustered by Coach, Independent Correlation Structure

Variable	**I**	**II**	**III**
Intercept	-1.872**	-2.137**	-2.155**
	(-2.676)	(-2.788)	(-2.864)
Season Win Percentage	-2.183**	-2.137**	-2.168**
	(-2.652)	(-2.563)	(-2.611)
Season Actual Win Percentage minus Expected Win Percentage	-6.796***	-6.805***	-6.755***
	(-5.585)	(-5.597)	(-5.488)
Number of Games Coached with Team	0.002	0.002	0.002
	(1.247)	(1.302)	(1.299)
(Number of Games Coached with Team)^2	-1.79e-06	-1.81e-06	-1.80e-06
	(-1.556)	(-1.577)	(-1.592)
Coach's Age in Years	0.030**	0.033**	0.032**
	(2.459)	(2.617)	(2.519)
Minority Coach Dummy Variable		0.210	0.177
		(0.997)	(0.847)
Former NBA Player Dummy Variable			1.756
			(0.894)
Estimated Scale Parameter	1.042	1.043	1.039
Number of Observations	595	595	595
Number of Dismissals	165	165	165

Coefficients for the variables in the logit model are shown with accompanying z-statistics. Statistical significance is noted with *-notation, where *** indicates statistical significance at the 1% level, ** indicates statistical significance at the 5% level, and * indicates statistical significance at the 10% level.

Appendix 3: Alternate Specifications of Logit Model: Inclusion of Metro Population and Ownership Change

Variable	I – Full Sample	II – Full Sample	III – Restricted Sample – Ownership Change Years Removed
Intercept	-1.582	-1.511	-1.263
	(-1.559)	(-1.484)	(-1.202)
Season Win Percentage	-2.425**	-2.458***	-2.547***
	(-2.548)	(-2.577)	(-2.613)
Season Actual Win Percentage minus Expected Win Percentage	-7.764***	-7.748***	-7.860***
	(-5.307)	(-5.287)	(-5.274)
Number of Games Coached with Team	0.002*	0.002*	0.002*
	(1.741)	(1.693)	(1.666)
(Number of Games Coached with Team)^2	-2.22e-06*	-2.26e-06*	-2.16e-06*
	(-1.947)	(-1.926)	(-1.855)
Coach's Age in Years	0.033	0.032**	0.026
	(2.066)	(2.059)	(1.610)
Minority Coach Dummy Variable	0.156	0.154	0.111
	(0.645)	(0.634)	(0.449)
Former NBA Player Dummy Variable	0.031	0.011	0.011
	(0.126)	(0.045)	(0.044)
Metro Population	-1.42e-08	-1.43e-08	-9.42e-09
	(-0.638)	(-0.640)	(-0.415)
Ownership Change Dummy Variable		-0.639	
		(-1.050)	
Season Dummies	YES	YES	YES
McFadden R^2	0.2038	0.2049	0.1956

Coefficients for the variables in the logit model are shown with accompanying z-statistics. Statistical significance is noted with *-notation, where *** indicates statistical significance at the 1% level, ** indicates statistical significance at the 5% level, and * indicates statistical significance at the 10% level.

Regional Inequality and Poverty Disparities between Social Groups in Rural India: A Decomposition Analysis

Snehasis Mondal[*] and Panchanan Das[†]

ABSTRACT

Poverty disparity between STs and non-ST rose substantially in rural India. This study decomposes poverty disparity to capture the reasons behind this rising poverty disparity. Unlike the conventional studies, this study uses two stage decomposition techniques to decompose poverty disparity. First stage decomposes poverty disparity into two components - WRD and BRD. Second decomposes WRD further into BID and BDD. This study reveals that marginal contribution of BRD is significant to poverty disparity between ST and non-SC/ST. However, BID has been found as the prime factor to the poverty disparity between the social groups.

1. INTRODUCTION

Scheduled Castes (SCs) and Scheduled Tribes (STs) were historically deprived in India. But reasons for their deprivation have been different. SCs face social discrimination but over time they have been able to integrate with mainstream society (Thorat & Mahamallik, 2007; Thorat & Newman, 2007). Alternatively, STs were geographically isolated and thus were delinked with mainstream development (Sen, 1992; Xaxa, 2001). Studies found that the STs have less access to irrigation facilities, market, roads and communication, electricity, health facilities and several other infrastructures relative to the non-STs (Chakraborty & Ghosh, 2000; Joshi, 1990; Kijima, 2006; Rao, 2003). Thus, reasons for impoverishment of socially marginalised groups SCs and STs are not homogeneous. SCs were backward because they were refused their basic rights of education, ownership of land, capital and several other economic characteristics (Thorat & Mahamallik, 2007; Thorat & Newman, 2007). Conversely, STs were destitute as the regions they resided were primitive, isolated and stagnant compared to the rest of India (Xaxa, 2001). Thus, regional differences have played a vital role in the division between STs and non-STs. Tribal population is disproportionately distributed across India. Majority of them resided in backward regions of India. Madhya Pradesh, Orissa, Jharkhand, Andhra Pradesh, North-Eastern regions of India, have largest proportions of STs considered as less developed regions in India. Few studies highlighted that geographical differences played a significant role to keep STs deep into poverty and thus responsible for prevalent poverty disparity[1] between STs and non-STs (Kijima, 2006; Sen, 1992; Xaxa, 2001). But there exists a lack of understanding about the contribution of geographical difference to the poverty disparity between STs and non-SC/ST.

[*] Assistant Professor, Vidyasagar College and Research Scholar, Department of Economics, University of Calcutta, Kolkata, West Bengal, India, email: snehasis.research@gmail.com

[†] Professor, Department of Economics, University of Calcutta, Kolkata, West Bengal, India

Mondal and Das (2021) decomposes poverty disparity between two social groups into ‘between groups income disparity’ and ‘between group distribution disparity’. Further, the changing disparity in poverty between social groups has been decomposed into - ‘difference in growth effect’ and ‘difference in distribution effect’ between the social groups. Mondal and Das (2021) reveals that the ‘growth effect’ among the SCs and non-SC/ST is significantly higher relative to that among the STs. Thus, it was concluded that growth in MPCE among the STs should be raised to the level of non-STs to counter this increasing poverty disparity between the STs and non-STs. This result however has a number of implications. The first, the growth in MPCE (Monthly Per Capita Consumption Expenditure)[2] among the STs is significantly lower, relative to the non-STs in majority of the regions in rural India. Consequently, on average, the growth among the STs is significantly lower relative to the non-STs. The second, the growth in MPCE among the regions, densely populated by the STs experienced a lower growth rate than the region where the STs are negligible in proportion. As a result, growth among the STs are less relative to the non-STs. Both of these propositions are logically valid. But the policy prescriptions should differ to either of these propositions. To counter the first specificities, social groups based on affirmative measures need more attention. On the other hand, to counter second specificity society requires to pay much attention on region based development. To point out empirically valid propositions, this study however decomposes poverty disparity into: ‘within regional disparity’ and ‘between region disparity’. Further, the marginal contribution of ‘within regional disparity’ to poverty disparity has been decomposed into ‘between groups income disparity’ and ‘between group distribution disparity’. Thus, this decomposition technique decomposes poverty disparity into three components - between region disparity (BRD), between groups income disparity (BID), and between group distribution disparity (BDD). BRD measures the contribution of regional disparity to poverty disparity. On the other hand, BID measures the contribution of mean income differential between groups to the poverty disparity and BDD measures the contribution of differential in distribution of income between the groups to poverty disparity. This decomposition can point-out the reasons for increasing poverty disparities between ST, SC and non-SC/ST in rural India. Thus, it will help the policy makers to design programmes to eliminate poverty disparities between social groups.

This work has been framed in the following manner. Section – 1 introduces the chapter. Section – 2 emphasises the decomposition technique developed in this paper. Section – 3 reveals application of the decomposition technique among social groups in rural India. Section – 4 analyses results. Section – 5 summarises and concludes this study.

2. SPATIAL DECOMPOSITION OF POVERTY DISPARITIES

Previous work by Mondal and Das (2021) decomposed poverty disparity between groups into between group income disparity and within group income disparity. The implicit assumption was there that either all regions are homogeneous in context of per-capita income of the regions or the proportion of social groups are uniformly distributed across the regions. Given the vast difference in per-capita income and incidence of poverty across the country, it is important to find out the contribution of the regional disparity on poverty disparity.

Attempts made in the past to decompose poverty over time into growth and distribution effect (Datt & Ravallion, 1992; Jain & Tendulkar, 1990; Kakwani & Subbarao, 1990; Maasoumi & Mahmoudi, 2013). These studies are dynamic in nature. On the other hand, few studies attempt to decompose poverty differential between regions into ‘within region disparity’ and ‘between region disparity’ using Shapley decomposition technique (Dhongde, 2004; Obayelu, 2014). Dhongde (2004), using the NSSO data for 1999-00, found that differences in incidence of poverty between states were largely contributed by the mean income disparity between the states. But all these conventional methods of decomposition did not decompose poverty disparity between groups ever. In this section, however, we decompose poverty disparity into three components - between region disparity (BRD), between groups income disparity (BID), and between group distribution disparity (BDD).

Poverty among a said group for a particular time depends on per-capita income and distribution among that group (Bourguignon, 2005). Thus, poverty disparity, referred to as log difference of poverty indices, between two groups in a region appears for either of the following reasons. Firstly, disparity in mean income between groups, i.e. one group comprises higher mean income relative to the other entitled with similar distribution in income among both the groups. Secondly, disparity in distribution between the groups, i.e. entitled with the same mean income one group has more equal distribution relative to the other. Thirdly, combination of both of the above. Thus, poverty disparity between two groups in a specified region can be decomposed into ‘between group income disparity’ and ‘between group distribution disparity’ (Mondal & Das, 2021). Poverty disparity between two groups for a country[3] has been decomposed within regional disparity and between regional disparity components in the following way.

2.1 NOTATIONS AND NEW DECOMPOSITION MEASURE

Let P_{ijt} represents poverty measure of group - i located in j^{th}region at time-t in such that $\sum_{j=1}^{n} w_{ijt}P_{ijt} = P_{it}$ where, $\sum_{j=1}^{n} w_{ijt} = 1$, w_{ijt} represents population share of group - i in jth region at time – t, and, P_{it} measures poverty among group - i at time - t. We further assume that $i = \{A, B\}, t = \{t, t+1\}, and, j = \{1,2, \dots \dots., n\}$. Poverty has been framed as the function of income, Lorenz function and poverty line. Further assuming that the poverty line is fixed over time and space, and poverty can be considered as a function of mean income and distribution (Datt & Ravallion, 1992; Kakwani, 2000; Maasoumi & Mahmoudi, 2013).

$$P = P\left(\frac{\mu^*}{z}, \frac{L^*}{z}\right) = P(\mu, L) \dots \dots ..(1)$$

Here, μ^* and L^* measures mean income and Lorenz function, $\mu \; and \; L$ denotes poverty line adjusted mean income and Lorenz function and z signifies the poverty line.

Existing decomposition techniques mentioned above tries to decompose poverty change over time into growth effect and distribution effect. This can be applied to decompose poverty disparities between two groups into ‘between group income disparity’ and ‘within group income disparity’ components. Apart from that, regional disparities will also have a significant impact on poverty disparities between social groups as the share of population of groups varies across the regions. Poverty disparity between two groups has been sketched as the log differences in poverty indices between two groups.

Poverty disparity between two groups (say A and B) in region-j at time-t is represented by

$$\Delta P(\mu_i, L_i)_{i \in A,B}$$

$$= logP(\mu_A, L_A) - logP(\mu_B, L_B) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (2)$$

Now, the above poverty disparities between groups can be decomposed into several components. Marginal contribution to poverty disparities of a said component refers to the variation in poverty disparities after adding the latter to the complement component set (Obayelu, 2014). To capture the marginal contribution of any component on some particular, two situations need to be compared - situation of the particular in presence of the component and the situation of the particular in absence of the component controlling all other relevant components. Thus, to capture the marginal contribution of regional disparity on poverty disparities, we must compare the poverty disparity between groups in presence of disparity between regions and in absence of disparity between regions. To introduce a situation where disparity between regions would be null i.e. no regional disparity, we consider all regions are associated with fixed real per-capita income i.e. real per-capita income among all regions are the same.

Let P_{ijt}^{μ} represents poverty measure among group - i located in j^{th}region at time-t considering that there exist constant mean per-capita income across regions. This has been done by scaling mean income of all regions equal. For this purpose, we multiply income of all household across all social groups by factor $\frac{\mu}{\mu_j}$ where, μ_j stands for per-capita income of jth region and μ represents composite per-capita income. Thus, $\sum_{j=1}^{n} w_{Aj} P_{Ajt}^{\mu} = P(\mu_A^{\tau}, L_A)$ measures the poverty among group – A in absence of regional disparity. This can be referred to as projected poverty among group-A in absence of regional disparity. Now, the difference between observed poverty among group-A, which prevailed in presence of regional disparity, and the projected poverty among group-A in absence of regional disparity reflect the marginal contribution disparity between regions on the poverty among group-A. Thus, $logP(\mu_A, L_A) - logP(\mu_A^{\tau}, L_A)$ measures the marginal contribution of disparity between regions of poverty among group-A. Similarly, $\sum_{j=1}^{n} w_{Bj} P_{Bjt}^{\mu} = P(\mu_B^{\tau}, L_B)$ presents the poverty among group – B in absence of regional disparity, if income of the people among this group scaled up by the multiplier $\frac{\mu}{\mu_j}$. And, $logP(\mu_B, L_B) - logP(\mu_B^{\tau}, L_B)$ measures the marginal contribution of disparity between regions of poverty among group-B.

Further assume Let $P_{.jt}^{\mu_j}$ represents poverty measure among in j^{th}region at time-t considering that per-capita income of all households of that region is equal to the mean per-capita income of the region. Thus, there exists no inequality within the region. For this purpose, we equate income of all household of a region to the mean per-capita of the said region (μ_j). Thus, $\sum_{j=1}^{n} w_{.j} P_{.jt}^{\mu_j} = P(\mu_j^{*}, L_j)$ measures projected poverty in absence of 'within regional disparity'. Consequently, $logP(\mu_B^{\tau}, L_B) - P(\mu_j^{*}, L_j)$ measures the marginal contribution of disparity within regions of poverty among group-B. And, $logP(\mu_A^\tau, L_A) - P(\mu_j^*, L_j)$ measures the marginal contribution of disparity within regions of poverty among group-A. Now combining the above two, $logP(\mu_A^\tau, L_A) - P(\mu_j^*, L_j) - logP(\mu_B^\tau, L_B) + P(\mu_j^*, L_j) = logP(\mu_A^\tau, L_A) - logP(\mu_B^\tau, L_B)$ presents marginal contribution within regional disparity to poverty disparity.

Now poverty disparity between group-A and group-B, as it has been framed in equation – 2, can be rewritten as follows

$$\begin{aligned} & logP(\mu_A, L_A) - logP(\mu_B, L_B) \\ &= logP(\mu_A, L_A) - logP(\mu_A^\tau, L_A) + logP(\mu_A^\tau, L_A) - logP(\mu_B^\tau, L_B) + logP(\mu_B^\tau, L_B) - logP(\mu_B, L_B) \\ &= logP(\mu_A, L_A) - logP(\mu_A^\tau, L_A) + logP(\mu_B^\tau, L_B) - logP(\mu_B, L_B) + logP(\mu_A^\tau, L_A) - logP(\mu_B^\tau, L_B) \\ &= BRD + WRD \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3) \end{aligned}$$

Here, $logP(\mu_A, L_A) - logP(\mu_A^\tau, L_A)$ measures marginal contribution of regional disparity to the poverty among group - A. And $logP(\mu_B^\tau, L_B) - logP(\mu_B, L_B)$ presents the marginal contribution of regional disparity to poverty among group – B. The equation 3 decomposes the poverty disparity between two groups viz, group – A and group – B, into between regional disparity (BRD) component and within regional disparity (WRD) component. Summing up the marginal contribution of regional disparity to the poverty among the groups we estimate the between regional disparity component. Thus, $logP(\mu_A, L_A) - logP(\mu_A^\tau, L_A) + logP(\mu_B^\tau, L_B) - logP(\mu_B, L_B)$ measures the contribution of regional disparity to the poverty disparity between groups. While marginal contribution of 'between regional disparity' netted out from the poverty disparity between groups, the residual expression can be termed as marginal contribution 'within regional disparity' to poverty disparity. $logP(\mu_A^\tau, L_A) - logP(\mu_B^\tau, L_B)$ presents marginal contribution within regional disparity to poverty disparity.

Thus, the above reasoning reveals that regional disparity is not the sole factor that accounts for the existing poverty disparity between groups. Poverty disparity may arise as the consequence of disparity among the regions i.e, disparity within the regions also. Contribution of WRD to poverty disparity appears because of differences in either group mean income, or within group distribution function, or both. Thus, the component, WRD, further has been decomposed into between group income disparity (BID) and between group distribution disparity (BDD) across social groups.

Further, assume $P_{Ajt}^{Bj.}$ denotes the mean adjusted poverty index among group-A at time-t in region-j. Here, per capita income of all households among group-A in region-j is multiplied by the factor $\frac{\mu}{\mu_j} * \frac{\mu_{Bjt}}{\mu_{Ajt}}$ in time period-t. μ_{Ajt} is the mean value of per capita income of all households in group-A in region-j at time point *t*, and μ_{Bjt} is the mean per capita income of group-B in the same region and same period. The mean adjusted poverty index for group-A measures the poverty among group – A considering distribution among A unchanged but mean per-capita income among group – A scaled to the mean per-capita income among group – B under a situation of null regional disparity. Similarly, $P_{Bjt}^{Aj.}$ stands for mean adjusted poverty index among group-B. It measures poverty among group – B considering distribution among A unchanged but mean per-capita income among group – B scaled to the mean per-capita income among group – A under a situation of null regional disparity.

$$\sum_{j=1}^{n} w_{Aj} P_{Ajt}^{Bj.} = P(\mu_A^{\tau B}, L_A) \qquad \& \qquad \sum_{j=1}^{n} w_{Bj} P_{Bjt}^{Aj.} = P(\mu_B^{\tau A}, L_B)$$

Now, $logP(\mu_A^{\tau}, L_A) - logP(\mu_A^{\tau B}, L_A)$ measures log difference of the poverty index due to the difference in group mean, keeping distribution among the group remains the same and with no regional disparity, when group-A is considered as the reference group. Similarly, $logP(\mu_B^{\tau A}, L_B) - logP(\mu_B^{\tau}, L_B)$ provides the same measure when group-B is considered as a reference group. Thus, poverty disparity between two groups due to differences in mean income is the average of these two components:

$$\frac{1}{2}[logP(\mu_A^{\tau}, L_A) - logP(\mu_A^{\tau B}, L_A) + \ logP(\mu_B^{\tau A}, L_B) - logP(\mu_B^{\tau}, L_B)]$$

The above expression reveals the marginal contribution of *between group income disparity (BID)* on poverty disparity. When per capita income of both groups is the same then the BID appears to be zero. A positive value of BID indicates that Group – A is possessed with less mean income relative to Group – B. On the other hand, the negative value of BID states that the Group – A is endowed with higher mean income relative to Group – B. Thus, deviation of its value from zero to either direction indicates income disparity between groups.

Apart from BID, distribution among the groups plays a significant role in explaining poverty disparity between groups. Marginal contribution of between group distribution disparity to the poverty disparity can be stated in the following manner.

$$\frac{1}{2}[logP(\mu_A^{\tau}, L_A) - \ logP(\mu_B^{\tau A}, L_B) + logP(\mu_A^{\tau B}, L_A) - logP(\mu_B^{\tau}, L_B)]$$

When the distribution of income of both the groups is the same BDD appears to be zero.

The desirable properties of decomposition measures are as follows

a. **Residual Free:** The properties of residual free and group-reversion consistency require followings. Let C be the index set for component of interest $\Delta^c D$ be the contribution of the component – c (such that c ε C) to poverty change. Decomposition measure $(C, \{\Delta^c D(i_t, j_t)\})$ will satisfy the property of residual free if and only if the following holds.

$$\Delta D(i_t, j_t) = \sum_{c} \Delta^c D(i_t, j_t)$$

b. **Group Reversion Consistency:** The decomposition $(C, \{\Delta^c D(i_t, j_t)\})$ will satisfy the property of group-reversion consistency if and only if the following holds.

$$\Delta^c D(i_t, j_t) = - \Delta^c D(j_t, i_t) \ , \forall \ C, t, i, j.$$

Our decomposition measure satisfies the property of residual free and group-reversion consistency. Group-reversion consistency requires that if we revert the order of the group, the reverse decomposition for a given time will produce the same decomposition results except that each component has the opposite sign.

3. DATA

This study is limited to the rural areas of India. This study is based on the survey data on household consumer expenditure, employment and unemployment situation, and periodic labour force conducted by the National Sample Survey Office (NSSO) in rural areas of India. This study covers 67 regions covering the rural areas of 26 states in India. The Union Territory has been excluded from this study. The household level data used in this study comes from the 38th, 43rd, 50th, 55th, 61st and 68th rounds quinquennial surveys on ‘Consumption Expenditure of Households’ conducted by the NSSO for the years 1983, 1987-88, 1993-94, 1999-2000, 2004-05 and 2011-12 respectively. The 68th round on Consumption Expenditure Survey is the ninth and the most recent quinquennial round for 2011-12. These surveys are a cross-section of a geographically distributed random sample of households. The survey period is one year for most of the quinquennial surveys. Generally, this period extended from July to June.

The survey in each round of this type is based on stratified multi-stage sampling. The census villages in the rural sector and urban frame survey blocks in the urban sector are the first stage sample units. The final stage ultimate sample units are households selected by simple random sample without replacement (SRSWOR) in both the sectors. The data set covers the geographical areas all over India, excepting few regions[4]. The cross-sectional survey is roughly representative of the national, state, and so-called “NSS region” level. Schedule 1.0 of these rounds collected consumption expenditure on different food and non-food items and some other characteristics, namely, age, sex and educational level, of every person within a household. The household consumer expenditure survey is the primary data source in estimating poverty in India.

NSS classifies social groups into four categories: Scheduled Tribe, Scheduled Caste, Other Backward Class, and Others consisting of the upper castes. The social groups in India are heterogeneous in nature with considerable intra-group variations. The survey data on consumption expenditure in NSS rounds are available at the household level. As there is no income survey in India, in this study we have used per monthly capita consumption expenditure as provided in the survey rounds as a proxy for income in measuring poverty. During the survey period the sample households are interviewed in two distinct recall periods - 30 days recall and 365 days recall[5]. Households are asked to reveal the expenditure on each consumable item during the last 30 days. Further, 365 days recall is used for selected items that come under the categories of clothing, footwear, durables, education and medical. Depending on these recall periods, three distinct estimates on MPCE - URP, MRP, and MMRP are presently available in NSSO-CES. Uniform Recall Period (URP) considers the consumption expenditure information collected using 30 days reference for all consumable items. Mixed Recall Period (MRP) takes into account the consumption expenditure data using 365 days recall for five least consumable items and 30 days recall for rest of the items. Modified Mixed Recall Period (MMRP) uses three distinct reference periods to collect the information on expenditure of the households. It uses 7 days recall for most frequent consumable items such as pan, tobacco, edible oil, egg, fish and meat, vegetable, fruits, spices and processed food; 365 days recall for five least frequent items; and, 30 days recall for rest of items. NSSO introduced MMRP from 2009-10. MMRP provides the most accurate estimates of MPCE for households followed by MRP. However, MMRP became available from 2009-10. Whereas, MRP based MPCE became available from 1993-94. But URP based MPCE is available for all quinquennial rounds of NSSO-CES, except in 1999-00. Thus, to maintain uniformity URP based MPCE has been used here.

Periodic Labour Force Surveys (PLFS) are being conducted by NSSO from 2017-18. These surveys collected consumption expenditure on different food and non-food items and some other characteristics, namely, age, sex and educational level, of every person within a household. The Labour Force Survey conducted in 2018-19 has been used here as it collected MPCE information of households along with several other endowment information at household and personal level. The Labour Force Survey conducted in 2018-19 covers the time period from July, 2018 to June, 2019.

4. EMPIRICAL FINDINGS

Poverty has been measured using Monthly Per Capita Consumption Expenditure (MPCE) of households. The Planning Commission published the estimates on the poverty line from 1973-74. Various Expert Committees have been formed to investigate the poverty line during 1973-74 to 2011-12. This study uses the Tendulkar Committee recommended poverty line to measure poverty as it covers a major time period of our analysis. The Planning Commission as the source of official poverty estimates was terminated in 2014. And reassembled as Niti Ayog since then. No official poverty estimates have been there since 2011-12. However, this research work uses the Consumer Price Index for Agricultural Labourer (CPIAL) to update the rural poverty line for 2017-18 to estimate poverty using Periodical Labour Force Survey (PLFS).

4.1 DISPARITIES IN POVERTY AND INCOME IN AGGREGATE

Poverty Gap Ratio (PGR) and Squired Poverty Gap (SPG) has been calculated among all social groups over the period of analysis using household level MPCE provided by various quinquennial rounds of NSSO Consumption Expenditure Surveys and Labour Force Survey. Table 1 reveals a declining trend of poverty between 1983 to 2011-12 among all social groups. However, the poverty situation worsens in 2017-18 relative to 2011-12 among all social groups. Studies argue that the demonetisation and poor implementation of Goods and Services Tax are responsible for economic slowdown and thus forced the poverty rates to increase (Mehrotra and Parida, 2020).

Table 1: Trend in poverty and poverty disparity among social groups

Poverty measure	PGR					SPG				
Social groups	1983	1993-94	2004-05	2011-12	2017-18	1983	1993-94	2004-05	2011-12	2017-18

STs	0.31	0.21	0.17	0.12	0.16	0.15	0.09	0.07	0.04	0.07
SCs	0.27	0.19	0.13	0.08	0.11	0.12	0.08	0.04	0.02	0.04
Non-SC/ST	0.18	0.13	0.08	0.04	0.08	0.08	0.05	0.03	0.01	0.03
$logP_{ST} - logP_{non\ SC/ST}$	0.24	0.23	0.33	0.43	0.30	0.29	0.27	0.42	0.54	0.35
$logP_{SC} - logP_{non\ SC/ST}$	0.17	0.18	0.21	0.23	0.14	0.21	0.21	0.25	0.26	0.14
$logP_{ST} - logP_{SC}$	0.06	0.05	0.11	0.20	0.16	0.09	0.06	0.17	0.28	0.20

Source: Authors' Calculation from different rounds of NSSO-CES and PLFS

Table – 1 further reveals that the hierarchy of social groups remains unchanged. STs comprise the highest rate of poverty followed by SCs. Despite the declining trend of poverty statistics among all social groups, poverty disparities between social groups have shown an increasing trend during the period of analysis. As it has been shown in the Table 1 that poverty disparities between ST and non-SC/ST, using PGR as poverty measure, rose significantly from 0.23 in 1993-94 to 0.38 in 2011-12. It expands further, from 0.27 during 1993-94 to 0.45 in 2011-12, when SPG has been taken into account. However, poverty disparities between ST and non-SC/ST, using both PGR and SPG, declined in the pre-reform period from 1983 to 1993-94. Thus, despite the decline in poverty disparities between ST and non-SC/ST in pre-reform periods there exists an increasing trend in poverty disparities between these groups in post-reform decades. However, during 2011-12 to 2017-18, it decreased from 0.38 to 0.3. But, in this period poverty rose among both the social groups. Studies argue that the demonetisation and poor implementation of Goods and Services Tax are responsible for economic slowdown and thus forced the poverty rates to increase (Mehrotra and Parida, 2021). However, the poverty estimates for 2017-18 is based on LFS data, which is not comparable with CES data collected by NSSO.

Poverty disparity between SC and non-SC/ST rose continuously till 2004-05 in trifling manner. But it decreased in the subsequent periods. Interestingly, poverty disparity rose significantly between two marginalised social groups – SC and ST as poverty declined at a higher rate among SC relative to that among ST. Statistics reveal that during the period of analysis poverty disparities between SC and non-SC/ST reduced but it rose significantly between ST and non-SC/ST. Further, between SC and ST, it rose significantly during 1983 to 2017-18. This is a major issue of concern among the policy makers. It shows that SCs reaped the benefits of affirmative measures in a better way relative to STs in rural India. Xaxa (2001) pointed out that the pattern of historical development among SCs and STs has been significantly different. SCs were socially excluded whereas ST were excluded geographically from the mainstream development (Xaxa, 2001). Studies pointed out that poverty is determined by average income and distribution of the society (Bourguignon, 2003). Given distribution, higher mean income forced the poverty to reduce. On the other hand, keeping the mean income the same, a better distribution helps to reduce poverty (Datt and Ravallion, 1992; Kakwani, 1993; Bourguignon, 2003). Table – 2 reveals the mean real MPCE, considering 2011-12 as base year and Delhi (rural) as base state, among different social groups over time. It shows that average income declined in 1993-94 from 1983 but thereafter it showed an increasing trend till 2011-12 among all social groups. This study considers income disparity between social groups as log differences in mean income between two social groups.

Table 2: Mean income gap between social groups[6]

	1983	1993-94	2004-05	2011-12	2017-18
ST	1026	1231	1334	1655	1452
SC	1067	1199	1395	1817	1541
non-SC/ST	1351	1573	1801	2315	1750
$\underline{Y}_{non-SC/ST} - \underline{Y}_{ST}$	325***	341***	466***	661***	299***
$\underline{Y}_{non-SC/ST} - \underline{Y}_{SC}$	283***	374***	405***	498***	210***
$\underline{Y}_{SC} - \underline{Y}_{ST}$	41***	-32	61***	162***	89***

Source: Authors' Calculation from different rounds of NSSO-CES and PLFS

*** significant at 99%

Table – 2 shows that mean income among all the social groups increased steadily during 1983 to 2011-12. However, it declined significantly in 2017-18 in 2017-18. It further reveals that the difference in mean income between SC and non-SC/ST rose during 1983 to 2011-12. However, this difference reduced to Rs 299 in 2017-18. But in 2017-18 mean income of all the social groups declined relative to 2011-12. It showed a divergence of mean income between SC and non-SC/ST. However, the divergence of mean income is much more between ST and non-SC/ST relative to that between SC and non-SC/ST. The difference in mean income between ST and non-SC/ST rose significantly from Rs 325 in 1983 to Rs 661 in 2011-12. But income disparity between ST and non-SC/ST rose significantly during the same period in rural India. Further, income disparity between two social groups – ST and SC rose drastically despite mean income among all social groups rose during 1993-94 to 2011-12. It reveals that the rate of growth in income is less among the ST relative to non-ST. Thus, growth in income might play a dominant role to increase poverty disparities between ST and non-ST. The above statistics reveals that SCs fared better in terms of poverty reduction and growth relative to STs. Therefore, catching up, which worked for SCs, does not work in case of STs. Kijima (2006) pointed out that geographical location differences partially affected the well-being differential between ST and non-ST. This study tries to examine the effect of geographical differences on poverty disparities between groups.

4.2. DISAGGREGATED DISPARITIES

Disparity at aggregate level is often misleading and not able to capture the presence of significant regional differences. Especially when ST population is heterogeneously distributed across the regions. Table 3 depicts the distribution of population of different social groups across the states in rural India. It reveals that Arunachal Pradesh, Meghalaya, Mizoram, and Nagaland are the states with a very large proportion of ST population relative to non-ST. On the contrary, the states like Uttar Pradesh, Bihar, Punjab, Haryana, Uttarakhand share a very insignificant proportion of ST population relative to non-ST. In between, the states in central India, such as Jharkhand, Chhattisgarh, Madhya Pradesh, and Odisha share a moderate proportion of ST population relative to the non-ST.

Table 3: Percentage of Population of Different Social Groups across the States in Rural India

Social Group	ST		SC		non-SC/ST	
State	2001	2011	2001	2011	2001	2011
Andhra Pradesh	8.4	9.3	18.4	19.2	73.2	71.5
Arunachal Pradesh	69.7	74.1	0.4	0	29.9	25.9
Assam	13.6	13.7	6.7	6.8	79.7	79.5
Bihar	1	1.4	16.4	16.6	82.6	82
Chhattisgarh	37.6	36.9	11.4	12.8	51	50.3
Gujarat	21.6	23.1	6.9	6.6	71.5	70.3
Haryana	0	0	21.4	22.5	78.6	77.5
Himachal Pradesh	4.3	6.1	25.6	26	70.1	67.9
Jammu & Kashmir	13.8	15.4	8.3	8.2	77.9	76.4
Jharkhand	31	31.4	12.4	12.6	56.6	56
Karnataka	8.4	9.2	18.4	20	73.2	70.8
Kerala	1.5	2.5	10.8	10.4	87.7	87.1
Madhya Pradesh	25.8	27.2	15.6	15.7	58.6	57.1

Maharashtra	13.4	14.6	10.9	12.2	75.7	73.2
Manipur	44.4	45.6	1.3	2.7	54.3	51.7
Meghalaya	90.2	90.1	0.4	0.5	9.4	9.4
Mizoram	96.3	96.6	0	0.1	3.7	3.3
Nagaland	93.7	92.8	0	0	6.3	7.2
Odisha	24.6	25.7	17.2	17.8	58.2	56.5
Punjab	0	0	33	37.5	67	62.5
Rajasthan	15.5	16.9	17.9	18.5	66.6	64.6
Sikkim	21.2	36.6	5	4.4	73.8	59
Tamil Nadu	1.6	1.8	23.8	25.5	74.6	72.7
Tripura	36.5	41.2	17.2	16.1	46.3	42.7
Uttarakhand	3.8	3.8	19.9	21.3	76.3	74.9
Uttar Pradesh	0.1	0.7	23.4	23	76.5	76.3
West Bengal	7.2	7.8	26.9	27.5	65.9	64.7

Source: Census of India, 2001 and 2011 (Excluding 3 sub-division of Senapati district of Manipur)

Thus, the north-eastern states share a significantly higher share of ST population relative to non-ST within the states. However, these north-eastern states such as Nagaland, Mizoram, Meghalaya, Tripura, Manipur, Sikkim, and Arunachal Pradesh, as has been depicted in Table 4, contributed only 10 percent of total ST population. Madhya Pradesh contributed the highest proportion of total ST population, followed by Maharashtra, Odisha, Rajasthan, Jharkhand, Gujarat, Chhattisgarh, Andhra Pradesh and West Bengal. These states contribute nearly 80 percent of the total ST population. Thus, the majority of STs concentrated in these eight states is located at the central-east and western part of India. However, SCs and non-SC/ST are not concentrated in such narrow regions. Except North-eastern states, which share the total population very marginally, SC and non-SC/ST have a significant proportion of their total population in each state. Each state has been segregated into a few NSS-Regions by NSSO during the survey. Thus, NSS-regions produce more disaggregated statistics relative to the states. In this chapter we hypothesise that regional disparity between NSS-regions have not contributed to the poverty disparity between social groups. Relative to the NSS-region, district level analysis would produce a much disaggregated picture of poverty disparity analysis. In this section we try to capture the regional disparity between the regions/districts as well as the poverty disparity between groups among the regions/districts.

Table 4: Proportion of ST & SC in Each State to National SC & ST Population (Rural)

	2001		2011	
	ST	SC	ST	SC
Andhra Pradesh	6.01	7.68	5.58	7.05
Arunachal Pradesh	0.78	0.00	0.84	0.00
Assam	4.08	1.17	3.91	1.19
Bihar	0.93	9.16	1.35	9.97
Chhattisgarh	8.10	1.43	7.71	1.63
Gujarat	8.88	1.64	8.55	1.48
Haryana	0.00	2.41	0.00	2.42
Himachal Pradesh	0.31	1.05	0.40	1.04
Jammu & Kashmir	1.36	0.48	1.50	0.49
Jharkhand	8.40	1.95	8.39	2.05
Karnataka	3.79	4.82	3.66	4.87
Kerala	0.45	1.92	0.46	1.18
Madhya Pradesh	14.80	5.20	15.22	5.37
Maharashtra	9.68	4.58	9.60	4.87
Manipur	0.91	0.02	0.84	0.03
Meghalaya	2.18	0.01	2.28	0.01
Mizoram	0.56	0.00	0.54	0.00
Nagaland	2.00	0.00	1.39	0.00

Odisha	9.95	4.04	9.59	4.04
Punjab	0.00	4.00	0.00	4.22
Rajasthan	8.69	5.82	9.27	6.20
Sikkim	0.13	0.02	0.18	0.01
Tamil Nadu	0.71	6.25	0.70	6.16
Tripura	1.25	0.34	1.19	0.28
Uttarakhand	0.31	0.94	0.28	0.97
Uttar Pradesh	0.12	23.17	1.10	23.19
West Bengal	5.35	11.67	5.17	11.11
ALL India	100	100	100	100

Source: Census of India, 2001 and 2011 (Excluding 3 sub-division of Senapati district of Manipur)

Table 5 depicts state-wise per-capita MPCE over time. It reveals that growth in mean per-capita MPCE varies across the states during 1983 to 2011-12. Kerala showed a significant growth rate during this period. Conversely, the North-eastern states such as Assam, Tripura, Manipur, Meghalaya, Sikkim etc. showed significantly lower rate of growth in per-capita MPCE during the period. As it has been mentioned earlier that the ST population majorly located at central-eastern and western portions comprises the states like Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Jharkhand, Gujarat, Chhattisgarh, Andhra Pradesh and West Bengal.

Table 5: State wise Real Mean MPCE in Rural India

	1983	1993-94	2004-05	2011-12	2017-18	Growth 1983 to 2011-12	Growth 1983 to 2017-18
Andhra Pradesh	1286	1447	1630	2200	1763	2.54	1.09
Assam	1233	1152	1340	1527	1819	0.85	1.40
Bihar	1045	1121	1159	1564	1447	1.77	1.13
Gujarat	1268	1341	1522	1974	1530	1.99	0.61

Haryana	1598	1608	2091	2321	1573	1.61	-0.05
Karnataka	1206	1238	1534	1920	1403	2.12	0.48
Kerala	1500	1693	2394	3238	2023	4.14	1.03
Madhya Pradesh	1262	1357	1314	1675	1521	1.17	0.60
Maharashtra	1124	1257	1432	1879	1373	2.40	0.65
Orissa	963	1197	1159	1534	1489	2.12	1.61
Punjab	1895	1818	1871	2426	2238	1.00	0.53
Rajasthan	1486	1456	1507	1920	1587	1.04	0.20
Tamil Nadu	1174	1399	1673	2085	1788	2.77	1.54
Tripura	1404	1499	1302	1673	2190	0.68	1.65
Uttar Pradesh	1297	1364	1510	1683	1532	1.06	0.53
West Bengal	1103	1422	1537	1753	1648	2.11	1.45
Jharkhand	1024	1148	1271	1471	1491	1.56	1.34
Chhattisgarh	1130	1214	1290	1426	1328	0.93	0.52
Uttaranchal	1964	1610	1648	2150	2291	0.34	0.49
Jammu&Kashmir	1340	1532	1829	2179	2077	2.24	1.62
Nagaland	-----	1399	1770	1605	1010	---	-----
Manipur	1155	1098	1253	1374	1466	0.68	0.79
Mizoram	1075	1502	1506	1467	1656	1.30	1.59
Meghalaya	1405	1484	1595	1756	1771	0.89	0.77
Sikkim	1498	1489	1735	1913	1632	0.99	0.26
Himachal Pradesh	1930	1659	1976	2645	1988	1.32	0.09

Arunachal Pradesh	----	1448	1777	2176	2035	---	------

Source: Authors' Calculation from different rounds of NSSO-CES and PLFS (note – base year 2011-12 & base state is Delhi)

Among these states Madhya Pradesh, Odisha, Rajasthan, Jharkhand, Chhattisgarh, which comprises nearly 50 percent of total ST population in rural India, showed an annual growth rate below 1 percent. Combining this with the low growth rate among the north-eastern states, it reveals that the region covering nearly 60-65 percent of ST population in rural India, experienced a very insignificant growth rate during 1983 to 2011-12. Thus, it shows that increasing income disparity between ST and non-ST at national level, to some extent this disparity is contributed by the disparity between the regions as those states share the majority of ST population, experienced insignificant growth rate during 1983 to 2011-12 relative to the rest of India.

However, disparity between the states is not the sole reason for prevalent income disparity between ST and non-ST at national level. Disparity between social groups among the states played a significant role in the prevalent income disparity between groups at national level. Income disparity between the social groups has been drawn as the logarithm difference of mean MPCE between the groups as has been depicted in Table 6. It reveals that disparity in mean income between ST and non-SC/ST is relatively lower in the northern-eastern states such as Mizoram, Manipur, Nagaland, Assam, Arunachal Pradesh, Meghalaya etc. Among these, Manipur, Nagaland, Meghalaya etc are the states where ST are majority in state population. However, these states cover only 10 to 15 percent of the total ST population. On the other hand, the states, such as Andhra Pradesh, Gujarat, Madhya Pradesh, Maharashtra, Odisha, Rajasthan, Chhattisgarh, and West Bengal, which comprises nearly 70 percent of total ST population, experienced significantly higher income disparity between ST and non-SC/ST. Only the state, Jharkhand experienced a lower income disparity relative to the above states. Surprisingly, it also reveals that income disparity between ST and non-SC/ST among these states reduced significantly during 1983 to 2011-12. Thus, it is found that income disparity between ST and non-SC/ST at national level significantly contributed to income disparity between ST and non-SC/ST among the states. But it has also captured two major points in the above analysis. Firstly, states covering the majority of ST population experienced insignificant growth rate during 1983 to 2011-12 relative to the rest of India. And secondly, Majority of these states experienced declining income disparity between ST and non-SC/ST during 1983 to 2011-12.

Table 6: State wise Income Disparity between Social Groups in Rural India

	1983			2011-12		
	non-SC/ST &	non-SC/ST &	SC & ST	non-SC/ST &	non-SC/ST &	SC & ST

	ST	SC		ST	SC	
Andhra Pradesh	0.11	0.09	0.03	0.08	0.07	0.01
Assam	0.02	0.02	0.00	0.02	-0.04	0.06
Bihar	0.04	0.12	-0.08	0.11	0.11	0.00
Gujarat	0.17	0.11	0.06	0.20	0.07	0.12
Haryana	0.02	0.15	-0.14	0.05	0.16	-0.12
Karnataka	0.17	0.14	0.03	0.15	0.09	0.06
Kerala	0.04	0.15	-0.11	0.26	0.21	0.06
Madhya Pradesh	0.19	0.15	0.04	0.19	0.12	0.07
Maharashtra	0.12	0.09	0.03	0.22	0.04	0.18
Orissa	0.19	0.12	0.07	0.17	0.12	0.05
Punjab	0.11	0.14	-0.03	0.17	0.18	-0.01
Rajasthan	0.17	0.12	0.05	0.14	0.07	0.07
Tamil Nadu	0.09	0.19	-0.10	0.08	0.09	-0.01
Tripura	0.08	0.16	-0.08	0.07	0.02	0.05
Uttar Pradesh	0.04	0.08	-0.05	-0.04	0.10	-0.14
West Bengal	0.11	0.07	0.04	0.13	0.04	0.09
Jharkhand	0.08	0.08	0.00	0.10	0.06	0.04
Chhattisgarh	0.10	0.07	0.03	0.09	0.04	0.05
Uttaranchal	0.02	0.19	-0.17	0.04	0.02	0.03
Jammu & Kashmir	0.13	0.09	0.05	0.05	-0.03	0.08
Nagaland	-	--	---	-0.21	-0.25	0.05
Manipur	0.06	0.01	0.05	0.04	0.10	-0.05

Mizoram	0.05	-0.06	0.11	-0.15	-0.25	0.11
Meghalaya	-0.05	0.14	-0.19	0.01	---	----
Sikkim	0.00	0.34	-0.34	0.02	0.10	-0.07
Himachal Pradesh	0.08	0.13	-0.05	0.16	0.11	0.05
Arunachal Pradesh	----	----	-----	-0.11	-0.34	0.24

Source: Authors' Calculation from different rounds of NSSO-CES

4.3 DECOMPOSITION OF POVERTY DISPARITY

The result of the decomposition of poverty disparities between groups in rural India has been decomposed within regions, and between regional disparity components as presented in Table – 7, 7a and &7b. Decomposition at State, NSS-Region as well as district level have been displayed in these tables. Decomposition at district level reveals a more disaggregated picture. Within the region disparity component further decomposed into income disparity and distribution disparity between groups within the region. Table – 7, 7a and 7b indicate that income disparity within a region plays a dominant role to explain poverty disparities between ST and non-SC/ST. It reveals that higher mean income among non-SC/ST relative to ST is the key factor of existence poverty disparity between ST and non-SC/ST.

Table 7: Decomposition of Poverty Disparity between different Social Groups at State level

Year	Poverty Gap Ratio				Squared Poverty Gap			
	WRD		BRD	PD	WRD		BRD	PD
Social groups	BID	BDD			BID	BDD		
Between STs and non-SC/ST								
1983	0.23	-0.03	0.03	0.23	0.29	-0.04	0.03	0.28
1987-88	0.27	-0.03	0.02	0.26	0.34	-0.03	0.03	0.34
1993-94	0.24	-0.04	0.02	0.23	0.31	-0.06	0.03	0.28

1999-00	0.37	-0.09	0.05	0.33	0.45	-0.11	0.08	0.41
2004-05	0.42	-0.15	0.06	0.33	0.51	-0.18	0.10	0.43
2011-12	0.50	-0.12	0.06	0.44	0.59	-0.12	0.08	0.55
2017-18	0.31	-0.03	0.02	0.29	036	-0.04	0.03	0.36
Between SCs and non-SC/ST								
1983	0.22	-0.04	0.01	0.17	0.28	-0.06	-0.01	0.21
1987-88	0.26	-0.07	-0.02	0.18	0.34	-0.10	-0.02	0.22
1993-94	0.28	-0.07	-0.02	0.19	0.36	-0.11	-0.02	0.23
1999-00	0.30	-0.08	-0.02	0.21	0.38	-0.11	-0.02	0.25
2004-05	0.35	-0.13	-0.01	0.21	0.44	-0.18	0.00	0.25
2011-12	0.39	-0.16	0.00	0.23	0.49	-0.23	0.00	0.26
2017-18	0.20	-0.05	-0.02	0.13	0.23	-0.07	-0.01	0.15
Between STs and SCs								
1983	0.01	0.01	0.03	0.06	0.01	0.02	0.04	0.08
1987-88	0.03	0.01	0.04	0.08	0.04	0.03	0.05	0.12
1993-94	-0.02	0.01	0.04	0.03	-0.02	0.02	0.05	0.04
1999-00	0.09	-0.03	0.07	0.12	0.11	-0.04	0.10	0.17
2004-05	0.10	-0.05	0.07	0.12	0.14	-0.06	0.10	0.18
2011-12	0.17	-0.02	0.05	0.21	0.21	0.00	0.07	0.29
2017-18	0.14	-0.01	0.03	0.16	0.16	0.00	0.04	0.20

Source: Authors' Calculation from different rounds of NSSO CES

Table 7 decomposition at state level considering both PGR and SPG. It reveals that poverty disparity between ST and non-ST rose significantly in the post reform period. Marginal contribution of income disparity (BID) component within the region plays the dominant role here. Distribution disparity component (BDD) on the other hand appears to be negative between ST and non-SC/ST. This indicates that income is well distributed among STs relative to non-SC/ST. Between regional disparity components of poverty disparity measure the extent of poverty disparity arises due to geographical difference. Positive value of the marginal contribution of the regional disparity component to poverty disparity signifies that, in general, STs resided in lower income regions relative to non-SC/ST. Surprisingly, the marginal contribution of regional disparity and income disparity within regions rose during 1983 to 2011-12. Thus, both increasing regional disparity and income disparity within regions between STs and non-SC/ST has been found the prime factor behind the increasing poverty disparity between ST and non-SC/ST. Moving from State level decomposition analysis to NSS-region level. Then again, from NSS-region level to district level analysis, this study found that marginal contribution of BRD appears to have a more prominent and significant effect on the poverty disparity between STs and non-SC/ST. Table 7a and 7b displays decomposition at NSS-Region as well as district level using PGR and SPG.

Table 7a: Decomposition of Poverty Disparity between different Social Groups using PGR

Year	Decomposition at NSS Regional level				Decomposition at District level			
	WRD		BRD	PD	WRD		BRD	PD
Social groups	BID	BDD			BID	BDD		
Between STs and non-SC/ST								
1983	0.18	-0.02	0.08	0.24	-	-	-	0.24
1987-88	0.28	-0.06	0.04	0.26	0.22	-0.03	0.07	0.26
1993-94	0.24	-0.07	0.06	0.23	-	-	-	0.23

1999-00	0.36	-0.12	0.09	0.33	0.21	0.02	0.11	0.33
2004-05	0.41	-0.17	0.09	0.33	0.21	0.02	0.11	0.33
2011-12	0.51	-0.21	0.13	0.43	0.27	0.01	0.15	0.43
2017-18	0.30	-0.07	0.07	0.30	0.14	0.05	0.11	0.30
Between SCs and non-SC/ST								
1983	0.22	-0.05	0.00	0.17	-	-	-	0.17
1987-88	0.27	-0.08	-0.01	0.18	0.27	-0.06	-0.03	0.18
1993-94	0.28	-0.08	-0.01	0.19	-	-	-	0.19
1999-00	0.30	-0.08	-0.02	0.21	0.27	-0.05	-0.02	0.21
2004-05	0.36	-0.14	-0.01	0.21	0.33	-0.10	-0.02	0.21
2011-12	0.41	-0.18	0.00	0.23	0.29	-0.06	-0.02	0.21
2017-18	0.22	-0.07	-0.01	0.14	0.23	-0.06	-0.03	0.14
Between STs and SCs								
1983	0.01	-0.01	0.06	0.06	-	-	-	0.06
1987-88	0.03	0.01	0.05	0.08	-0.04	0.02	0.10	0.08
1993-94	-0.02	-0.01	0.06	0.03	-	-	-	0.03
1999-00	0.06	-0.05	0.11	0.12	-0.05	0.04	0.13	0.12
2004-05	0.07	-0.06	0.11	0.12	-0.08	0.07	0.13	0.12
2011-12	0.16	-0.08	0.12	0.20	-0.18	0.19	0.19	0.20
2017-18	0.10	-0.01	0.08	0.17	-0.02	0.05	0.14	0.17

Source: Authors' Calculation from different rounds of NSSO CES

** District level information are not available for year 1983 and 1993-94 in NSSO-CES data

Table 7a and 7b reveals that poverty disparity between two marginalised social groups ST and SC rose significantly during 1993-94 to 2011-12 i.e., in the post-reform era. Marginal contribution of regional disparity has been found the prime factor to poverty disparity. Further, this marginal contribution rose substantially during 1983 to 2011-12. Marginal contribution of BID appeared prominent to poverty disparity between ST and SC since 1993-94 at NSS-regional level. Since then, it started climbing up significantly such that in 2004-05 both the components – BRD and BID shared equal contribution. Marginal contribution of BDD to poverty disparity between ST and SC appeared negative since 1983. It reveals distribution among the SC appeared to be more skewed relative to the ST. At district level, which displays a more disaggregated picture, reveals that poverty disparity between two marginalised social groups majorly contributed by the BRD.

Table 7b: Decomposition of Poverty Disparity between different Social Groups using SPG

Year	Decomposition at NSS Regional level				Decomposition at District level			
	WRD		BRD	PD	WRD		BRD	PD
Social groups	BID	BDD			BID	BDD		
Between STs and non-SC/ST								
1983	0.28	-0.08	0.09	0.29	-	-	-	0.29
1987-88	0.35	-0.07	0.06	0.34	0.29	-0.05	0.10	0.34
1993-94	0.32	-0.11	0.07	0.27	-	-	-	0.27
1999-00	0.45	-0.15	0.11	0.41	0.28	0.00	0.13	0.41
2004-05	0.52	-0.22	0.11	0.42	0.28	-0.01	0.15	0.42
2011-12	0.62	-0.21	0.13	0.54	0.54	-0.17	0.17	0.54

2017-18	0.37	-0.09	0.07	0.35	0.19	0.05	0.11	0.35
Between SCs and non-SC/ST								
1983	0.27	-0.07	0.01	0.21	-	-	-	0.21
1987-88	0.34	-0.10	-0.02	0.22	0.34	-0.08	-0.03	0.22
1993-94	0.35	-0.11	-0.01	0.23	-	-	-	0.23
1999-00	0.37	-0.11	-0.02	0.25	0.33	-0.07	-0.01	0.25
2004-05	0.45	-0.19	-0.01	0.25	0.40	-0.15	0.01	0.25
2011-12	0.48	-0.22	0.01	0.26	0.30	-0.03	-0.01	0.26
2017-18	0.25	-0.09	-0.01	0.15	0.28	-0.09	-0.04	0.15
Between STs and SCs								
1983	0.01	-0.01	0.08	0.08	-	-	-	0.08
1987-88	0.04	0.01	0.08	0.13	-0.03	0.03	0.13	0.13
1993-94	-0.03	-0.01	0.08	0.08	-	-	-	0.08
1999-00	0.09	-0.05	0.13	0.17	-0.03	0.05	0.15	0.17
2004-05	0.11	-0.06	0.12	0.17	-0.06	0.08	0.15	0.17
2011-12	0.21	-0.06	0.13	0.28	-0.20	0.26	0.22	0.28
2017-18	0.11	0.00	0.09	0.20	-0.01	0.06	0.15	0.20

Source: Authors' Calculation from different rounds of NSSO CES

** District level information are not available for year 1983 and 1993-94 in NSSO-CES data

Conversely, poverty disparity between SC and non-SC/ST majorly contributed by BID since 1983. Marginal contribution of BRD appeared to be negligible here. However, the marginal contribution of BRD appeared substantial to the poverty disparity between ST and non-ST (both non-SC/ST and SC). Marginal contribution of BDD to poverty disparity between SC and non-SC/ST appeared negative since 1983. It

reveals distribution among the non-SC/ST appeared to be more skewed relative to the SC.
Marginal contribution of BID component plays the major role in poverty disparity between SC/ST and non-SC/ST. Contribution of regional disparity to poverty disparity is negligible between SC and non-SC/ST. But it contributes significantly to poverty disparity between ST and non-SC/ST. Thus, poverty disparity between SC and non-SC/ST entirely contributed by the BID component. Whereas, poverty disparity between ST and non-SC/ST contributed by both BID component and between regional disparity component. Major reason behind this is that although SC and non-SC/ST segregated socially, they were not segregated geographically. On the other hand, ST were segregated geographically from non-ST. Poverty disparity between two marginalised groups – ST and SC rose significantly over time. It dominantly contributed to regional disparity between SC and ST.

Poverty disparity between the social groups declined significantly in 2017-18 relative to 2011-12. This could have been a significant achievement. But, decline in poverty between social groups in 2017-18 is associated with significant rise in poverty among all social groups. Whereas during 1983 to 2011-12 poverty among all social groups declined steadily. Researchers argued that this rise in poverty is the effect of implementation of GST and distortion in monetary circulation *(Note-Bandi)*, which significantly affected the unorganised sectors adversely (Melhotra and Farida, 2021). This period experienced a rise in poverty at a higher rate among the non-SC/ST relative to the SCs and STs. Consequently, poverty disparity reduced between the social groups.

5. SUMMARY AND CONCLUSION

To understand the changes in poverty differences between social groups with heterogeneous socio-economic characters we decompose poverty disparity by applying an alternative methodology developed in this study. The empirical findings in this study are not similar to those available in the empirical literature on poverty in India. While some studies observed a converging trend in poverty across the social groups, we observe a diverging trend particularly between the tribes and non-tribes during the high growth phase of the post-reform period. In some studies, STs and SCs are taken together to form socially marginalised groups. In this study we observe that these socially vulnerable groups were affected in dissimilar manner during the post reform development in India. Poverty disparity between SCs and STs increased significantly during this period. This study finds that the incidence of poverty among the tribes declined at a lower rate than for other social groups. This study observes that poverty disparity between STs and non-SC/ST is increasing mainly because the rate of decline in poverty is higher among the upper castes relative to the tribal people.

As mentioned earlier, BDD measures the extent of poverty disparities that arise due to arising inequality among one group relative to another. Contrary, BID reveals the extent of poverty disparities arise due to increasing mean income of richer groups relative to poorer. Results reveal that BID is the prime factor of increasing poverty disparities between ST and non-ST. Hence, differential income between ST and non-SC/ST has been found to be the major cause of prevalent poverty disparity between these groups. However, increasing marginal contribution of BID indicates that SCs and non-SC/ST experienced higher levels of growth relative to STs in the post-reform era. Considering the fact that STs were isolated geographically, this paper tries to point out whether regional disparities do have any significant role in increasing poverty disparities between ST and Non-ST. This study found that marginal contribution of regional disparity is significant to poverty disparity between ST and non-SC/ST. Whereas, it appears negligible to the poverty disparity between SC and non-SC/ST.

Despite having several affirmative government actions targeted for the socially marginalised section the disparity in poverty between the upper castes and tribes has increased significantly. There is a popular belief that the benefits of the affirmative actions initiated and implemented by the government have been appropriated mostly by the relatively well-off households within the marginalised groups like STs or SCs. This analysis nullified this belief. We found that significantly lower growth among STs relative to the growth rate among Upper castes and SCs is the prime cause of increasing poverty disparities between STs and non-ST. Unlike SCs and non-SC/ST, STs resided in relatively backward regions in rural India. Considering it, this study reveals that along with affirmative measures like reservation in Government jobs, educational institutions, different poverty alleviation schemes such as MGNREGA etc, the policy makers should consider the development of ST populated regions to eliminate poverty disparities between social groups in general and between ST and non-SC/ST in specific.

ENDNOTES

1. Poverty Disparity between social groups has been considered the log difference of poverty indices between the social groups. Our earlier work showed that the log poverty difference between groups is superior to the normal difference for application. For detail, please look at Mondal and Das (2021)
2. MPCE has been used as a proxy of income in this study as there is no household level information on income available in India.
3. Country has been considered as a union of numbers of regions.
4. (i)Leh (Ladakh) and Kargil districts of Jammu & Kashmir, (ii) interior villages of Nagaland situated beyond five kilometres of the bus route and (iii) villages in Andaman and Nicobar Islands which remain inaccessible throughout the year.
5. During the 66th and 68th round of NSSO-CES 7 days recall has been used for most frequently consumable items.
6. MPCE has been considered as the proxy of Income. It has been measured in rupees with base year 2011-12 and base state is Delhi.

REFERENCES

Anand, I., & Azad, R. (2019). India’s Slowdown: What It Is and What can be Done. *Economic and Political Weekly 54*(41).

Bourguignon, F. (2005). The Effect of Economic Growth on Social Structures. *Handbook of Economic Growth* 1(Part B) 1701–1747.

Chakraborty, G., & Ghosh, P. K. (2000). *Human Development Profile of Scheduled Castes and Tribes in Selected States: A Bench Mark Survey* (No. 04). National Council of Applied Economic Research.

Datt, G., & Ravallion, M. (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s. *Journal of Development Economics*, *38*(2), 275–295.

Dhongde, S. (2004). Decomposing Spatial Differences in Poverty in India. *WIDER Working Paper Series*, *2004/53*.

Jain, L. R., & Tendulkar, S. D. (1990). Role of Growth and Distribution in the Observed Change in Headcount Ratio Measure of Poverty: A Decomposition Exercise for India. *Indian Economic Review*, *25*(2), 165–205.

Joshi, Y. G. (1990). *Development in Overexploited Tribal Regions*. Inter-India Publications.

Kakwani, N., & Subbarao, K. (1990). Rural poverty in India, 1973-86. *Policy Research Working Paper Series*, Article 526.

Kijima, Y. (2006). Caste and Tribe Inequality: Evidence from India, 1983–1999. *Economic Development and Cultural Change*, *54*(2), 369–404.

Maasoumi, E., & Mahmoudi, V. (2013). Robust growth-equity decomposition of change in poverty: The case of Iran (2000–2009). *The Quarterly Review of Economics and Finance*, *53*(3), 268–276.

Mehrotra, S., & Parida, J. (2021, August 3). Poverty in India is on the rise again. *The Hindu*. https://www.thehindu.com/opinion/lead/poverty-in-india-is-on-the-rise-again/article35709263.ece

Mondal, S., & Das, P. (2021). Decomposition of poverty disparity across social groups in rural India. *Journal of Income & Wealth (The*, *43*(1 & 2), 75–89.

Obayelu, O. A. (2014). Spatial Decomposition of Poverty in Rural Nigeria: Shapley Decomposition Approach. *Journal of Economics and Sustainable Development*, *5*(27), 169.

Rao, N. (2003). Jharkhand: Vision 2010. *Economic and Political Weekly*, *38*(18).

Sen, R. (1992). Tribal Policy of India. *Indian Anthropologist*, *22*(2), 77–89.

Thorat, S., & Mahamallik, M. (2007). *Chronic Poverty and Socially Disadvantaged Groups: Analysis of Causes and Remedies* (SSRN Scholarly Paper No. 1756784).

Thorat, S., & Newman, K. S. (2007). Caste and Economic Discrimination: Causes, Consequences and Remedies. *Economic and Political Weekly*, *42*(41), 4121–4124.

Xaxa, V. (2001). Protective Discrimination: Why Scheduled Tribes Lag Behind Scheduled Castes. *Economic and Political Weekly*, *36*(29).

Open Source Software Licensing and Developer Participation

Vidya Atal* and Kameshwari Shankar†

Abstract

This paper looks at incentives for developer participation under different open source software (OSS) licenses We look at two types of OSS licenses - (1) restrictive licenses that force all modifications to be released under the same open source license, and (2) non-restrictive licenses that allow contributions on the open source platform to be made proprietary by developers. We develop a model of effort provision from software developers under each kind of license and compare the outcome. We identify two motives governing developer effort -- user motive to improve surplus as a consumer of the software, and revenue motive as a potential proprietor of the new software. The results of our analysis explain empirically observed differences in developer participation across open source licenses. Our model also allows us to compare welfare across licenses.

1. Introduction

Several empirical research papers have observed that developer participation in open source software (OSS) depends on the specific type of open source license adopted. There is a wide range of OSS licenses that provide varying degrees of freedom to developers on how to modify, combine or attribute contributions across open source platforms. The most restrictive license, the GPL (or GNU Public License) requires all developers to not only make its source code public, but also requires all derivative works to be released under the same open licensing terms. On the other hand, other less restrictive licenses provide the source code under open licensing terms while allowing developers to make their contributions on the platform proprietary. A well-known example of the latter is the Apache license used by Google's Android operating system on mobile devices.

Papers that have empirically tested the relationship between OSS license and software development generally find that restrictive licenses are associated with lower levels of participation and performance compared to non-restrictive licenses.[1]

In this paper, we provide a preliminary theoretical framework to describe the relationship between developers' effort provision and OSS\ licensing that would explain these findings. Our analysis also allows us to compare social welfare associated with different licenses. We show that while restrictive licenses may be associated with lower effort provision than non-restrictive licenses, they can sometimes be more efficient. We argue that a non-restrictive license leads to over-provision of developer effort relative to what

* Department of Economics, Feliciano School of Business, Montclair State University, USA, Email: atalv@montclair.edu
† Department of Economics and Business, Colin Powell School, CCNY, USA, Email: kshankar@ccny.cuny.edu

is efficient. Hence, the empirical finding that less restrictive licenses lead to more developer participation does not necessarily imply greater optimality. Finally, endogenizing license choice, we show that a restrictive license is adopted in platforms targeted towards developers whereas non-restrictive licenses tend to be adopted in projects with a large number of users but relatively few developers.

We build a model of software development where user-developers independently exert costly effort into creating applications over a software platform under a given license. There are two motives that govern developer participation in an OSS platform: 1) user motive -- as a user, developers would like to enhance their consumer surplus by generating a high quality software; and 2) revenue motive -- when the license allows applications developed on the platform to be made proprietary, developers can appropriate revenues if they happen to generate successful applications with market value. Participation incentivized through the user motive generates positive externalities to other users and developers and hence is subject to the classic public good problem of under-provision. On the other hand, the revenue motive creates inefficient incentives to over-participate in order to win the proprietary prize. The trade-off between these opposing externalities then determines the optimality of each license.

Our paper is organized as follows. Section 2 briefly discusses existing literature on OSS development and highlights our contribution to it. Section 3 describes our model of OSS development under restrictive and non-restrictive OSS\ licenses and compares them to the optimal first-best license. Section 4 concludes and provides avenues for future research in this area. Proofs of all propositions are provided in the appendix.

2. Related Literature

Our paper contributes to two broad areas of research in OSS development - first is the motivation for OSS participation by software developers; and the second is the role of licensing in OSS development.

The question of what motivates software developers to participate in OSS development has intrigued researchers since its inception. By definition, open source software limits the use of intellectual property rights by software developers. This precludes an explicit monetary mechanism to compensate OSS developers for their effort. What then motivates programmers to develop open source applications? The literature on OSS has provided a few different explanations for this phenomenon. The first is an intrinsic motivation driven by ideology or altruism. Open source software traces its origins to the “free software movement” started by Richard Stallman, a computer programmer and hacker who forcefully pushed the philosophy that all software development should be open and free for anyone to develop, share and modify. Stallman and other prominent free software advocates have had a strong following among many software developers. At least the early years of OSS progress can be attributed to a sense of free creative expression as well as an obligation to the community of “hackers” among developers pursuing the goal of free software laid out by Stallman's ideology. Lakhani and Wolf (2005), with the help of survey data, find that enjoyment and intellectual stimulation as well as an obligation to the OSS\ community are strong motivators of participation among non-paid developers. Eric Raymond, another prominent free software advocate, argues that Linux hackers are driven by reputational concerns within the community of developers and ego satisfaction from solving complex programming problems.

However, many economists have pointed out that it is difficult to explain the long-term persistence and success of several open source projects simply on the basis of psychological or ideological motivations. Roberts et al. (2006) find that intrinsic motivations such as enjoyment are not positively correlated with higher participation. Lerner and Tirole (2002) express skepticism over the claim that altruism sustains OSS development. They point out that altruism has not played a significant role in the development of other industries and hence its over-sized impact in OSS needs an explanation.

From the perspective of an economist, then, we can turn to other economic or utilitarian motives for participation. The literature in economics has examined two primary incentives: 1) participation in OSS for improving the quality of the software for their own use; and 2) job market signaling, where developers signal their ability to potential employers through their publicly available work on OSS. In the first incentive, developers are also users of the software and hence have an incentive to contribute to its development for their own individual or business needs. Johnson (2002) follows this line of reasoning to develop a voluntary contributions model of OSS as a public good. A number of papers have also found evidence for the second motive. Fershtman and Gandal (2007), in a large dataset of programmer participation in OSS\ projects, find evidence for reputational and career incentives for participation. Similarly, Hann et al. (2013) look at Apache OSS\ projects which has a merit-based ranking of programmers. They find that a higher ranking within the Apache ranking is associated with a wage increase for the programmer.

Atal and Shankar (2015) argue that the motive for participating in OSS depends on the stage of development. Reputational incentives drive developer effort in early stages when there can be considerable gains in status from breakthrough contributions. On the other hand, maximizing user value becomes an important objective in later stages as the software is adopted by more and more end-users.

The current paper follows the Johnson (2002) framework of looking at OSS developers as users of the software. In addition, unlike previous models of motivations, we consider the role of software licensing in affecting the motivation for developer participation in an OSS project.

A number of recent papers have studied different aspects of software licensing as it relates to developer participation and OSS\ success. Lerner and Tirole (2005) look at data from a large OSS project database to correlate license type with project characteristics. Sen et al. (2008) find that OSS with different licenses attract differently motivated developers. Comino et al. (2007) show that OSS projects with highly restrictive licenses have a lower probability of reaching a mature stage of development. Fershtman and Gandal (2007) find that output per developer is lower under more restrictive licenses.

Much of the existing research on licenses has been confined to empirically examining the correlation between OSS licensing and software outcomes. To our knowledge, there are no structural models of licensing and participation that would explain observed empirical regularities and shed light on the mechanism by which license choice affects developer participation and project success. Our paper fills this gap to provide a theoretical basis for the observed findings and to examine efficiency trade-offs in each license.

3. Model and Analysis

We consider a model of development for an OSS platform with $N > 1$ user-developers who contribute to the value of the platform by exerting effort and use the platform that is created by the combined contributions of all N developers.[2] Apart from the N developers, the platform is also used by $M \geq 0$ users who are not developers. These users simply consume the final platform at market price without contributing to its development.

The platform can have two types of OSS licenses -- restrictive (r) and non-restrictive (nr). Under a restrictive license, the platform must stay open and no developer can make it proprietary, while under a non-restrictive license, one of the developers can become the proprietor of the platform and sell it to other users to generate positive revenues.

Given the software license, each developer i exerts effort $e_i \geq 0$ to develop the platform. The cost of effort is given by $C(e) = \frac{1}{\beta}e^{\beta}$. We assume that $\beta > 1$ so that the cost function is convex. Development of the platform occurs in the following stages. First, a license is chosen for the platform - r or nr. Next, the N developers simultaneously choose their effort, e, on the platform. After all developers exert effort, a critical developer k is realized through a stochastic process. The contribution made by the critical developer is essential to the platform in the sense that the platform cannot create any user-value without this developer's code. The identity of the developer k who generates the critical contribution is determined through a stochastic process that depends on the effort invested by the developer. Specifically, we define the stochastic variable, $y_i = e_i\varepsilon_i$ where $\varepsilon_i \sim U[0,1]$ is independently and identically distributed for each developer. The critical developer is then the one who generates the highest value of y_i.

The value of the platform, v, is the sum of every developer's contribution as long as it includes developer k's contribution. If $\mathcal{N} \subseteq \{1,2 \dots N\}$ is a subset of the N developers who exert effort on the platform, then the value of the platform is:

$$v = \begin{cases} \sum_{i \in \mathcal{N}} e_i \;\; if\; k \in \mathcal{N}, \\ 0 \; otherwise. \end{cases}$$

Before proceeding with the analysis, let us discuss the role of the critical developer in our model. The emergence of a critical developer in the process of OSS development in our model allows us to have a concrete proprietor selection process under a non-restrictive license. If the OSS platform adopts a nr license in the first stage, any of the N developers in stage 2 can make their contribution proprietary under the terms of the license. We assume, however, that there is a single developer, namely the critical developer k, who can effectively appropriate the value of her effort by making her contribution proprietary. This is because, hers is the only contribution that has stand-alone value in the platform. Every other developer's contribution has zero value on its own if it is not combined with the code contributed by developer k. One can of course assume other selection criteria that may allow for multiple proprietors under a nr license. While our analysis is robust to changes in this selection process, the current set-up with a single proprietor is the most tractable as it allows us to illustrate the trade-offs with the nr license most clearly.

A significant, and in our view realistic, assumption about proprietor selection in our model is that both “luck” and “hard work” play a role. There is a non-zero probability that any developer can make the crucial innovation even if other developers put in more effort. At the same time, as the analysis in the next section will show, the probability of being the critical developer is increasing in effort, so that a developer who exerts greater effort is more likely to become proprietor under a nr license.

Given this set-up, the payoff to the developer has two possible components. First, under a non-restrictive license where the developer can make the platform proprietary, she may receive revenue from selling her application if she turns out to be the critical developer k. Second, as a user of the application on the platform, developers receive user-value from the final product. For any developer i, let us denote revenue from proprietary sales by π_i and user value from consumption by u_i. The surplus to developer i, denoted by S_i, is her payoff minus effort cost, i.e.,

$$S_i = \pi_i + u_i - C(e_i).$$

3.1. First Best Efficient Effort

To begin with, let us see how a social planner chooses effort for each developer. In this first-best efficient outcome, the social planner maximizes the total value for all N developers minus the total cost of effort from these N developers. If the social planner chooses e_o for each developer, platform value is $\sum_{i=1}^{N} e_o = Ne_o$. This is consumed by $(M+N)$ users. Thus, the total surplus from effort to be maximized is:

$$max_{e_o}\left[(M+N)Ne_o - N\frac{1}{\beta}e_o^{\beta}\right].$$

Taking the first order condition (which is necessary and sufficient given convex costs), first-best level of effort provision solves $(M+N) - Ne_o^{\beta-1} = 0$. Hence,

Equation 1

$$e_o^* = (M+N)^{\frac{1}{\beta-1}}.$$

In the next two subsections, we derive the equilibrium effort chosen by developers under each type of OSS license - restrictive and non-restrictive and compare it to e_o.

3.2. Restrictive OSS License

With a restrictive license, no single developer can make the platform proprietary. In particular, critical developer k is forced to keep her contributions open and hence cannot appropriate any revenue from the output she generates. Hence, developer surplus from effort only arises from the user-value created by the platform for its developers. Thus, surplus to developer i, given the effort of the remaining $(N-1)$ developers, is:

$$S_i(e_i; e_{j\neq i}) = u_i(e_i; e_{j\neq i}) - \frac{1}{\beta} e_i^{\beta} = \left(e_i + \sum_{j\neq i} e_j\right) - \frac{1}{\beta} e_i^{\beta}.$$

The surplus function is concave and hence the effort level that maximizes developer surplus is $e_r^* = 1$. Proposition 1 below compares effort provision under a restrictive OSS license to the optimal effort, e_o^*.

<u>*Proposition 1*</u>

The restrictive OSS license leads to lower participation from developers than what is efficient, i.e., $e_r^* < e_o^*$.

The restrictive OSS license under-provides effort relative to the optimal level due to the public good characteristics of OSS under this license. Software value to users is determined by aggregate developer effort and all users can simultaneously derive value from the platform so that the value created is non-rival. At the same time, the mandatory openness prevents the exclusion of any user from consuming the platform. As a result, value generation is less than optimal under this license.

3.3. Non-Restrictive OSS License

Next, we look at effort provision under a non-restrictive license. Under our set-up, only developer k who realizes the critical innovation to the platform, can appropriate value with a proprietary license. The remaining developers with non-critical contributions do not create anything with stand-alone user-value. Thus, critical developer, k, makes her contribution proprietary and appropriates all value from the platform by setting a monopoly price that equals the platform value, v. Hence, $\pi_k = (M+N)v$ and $S_k = (M+N)v - \frac{1}{\beta} e_k^{\beta}$; whereas the other developers who have to buy the platform from the proprietor, receive no user surplus but have already incurred the cost of effort by contributing to the platform. As a result, all other developers earn negative surplus, i.e., $S_i = 0 - \frac{1}{\beta} e_k^{\beta}$, for all $i \neq k$. In direct contrast to the restrictive license, developer effort under a non-restrictive license is entirely driven by revenue incentives.

Since the identity of developer k is unknown ex ante at the time of effort provision, the expected surplus to the developer from exerting effort e_i is:

$$E\left[S_i(e_i; e_{j\neq i})\right] = \Pr[i=k]\,(M+N)\left(e_i + \sum_{j\neq i} e_j\right) - \frac{1}{\beta} e_i^{\beta}.$$

Restricting our attention to symmetric equilibria, we derive the probability density function for k. Since the critical developer is the one with the highest y_i, the probability that developer i is developer

k is, $\Pr(i = k) = \Pr\,(y_i > y_j, for\ all\ j \neq i)$, where developer i chooses e_i and all other developers choose $e_j = e_{nr}$. Deriving the density function using the uniform distribution ε_i gives us the following:[3]

Equation 2

$$\Pr(i = k) = \begin{cases} \dfrac{1}{N}\left(\dfrac{e_i}{e_{nr}}\right)^{N-1} \ if\ e_i < e_{nr}, \\ \dfrac{1}{N}\left[N - (N-1)\left(\dfrac{e_{nr}}{e_i}\right)\right] \ if\ e_i \geq e_{nr}. \end{cases}$$

The expected surplus to developer i is then, given by

$$E[S_i(e_i; e_{nr})] = \begin{cases} \dfrac{(M+N)}{N}\left(\dfrac{e_i}{e_{nr}}\right)^{N-1}[e_i + (N-1)e_{nr}] - \dfrac{1}{\beta}e_i^{\beta} \ if\ e_i < e_{nr}, \\ \dfrac{(M+N)}{N}\left[N - (N-1)\left(\dfrac{e_{nr}}{e_i}\right)\right][e_i + (N-1)e_{nr}] - \dfrac{1}{\beta}e_i^{\beta} \ if\ e_i \geq e_{nr}. \end{cases}$$

Under appropriate assumptions on β and N, there is a symmetric equilibrium with positive effort.[4] Solving the first order condition and setting $e_i = e_{nr}$, we get positive equilibrium effort under a nr license as:

Equation 3

$$e_{nr}^{*} = \left[(M+N)\left(N - 1 + \frac{1}{N}\right)\right]^{\frac{1}{\beta-1}}.$$

The following proposition compares effort provision and software value across the two licenses and with the efficient outcome.

<u>*Proposition 2*</u>

Relative to the first-best outcome, a restrictive OSS\ license under-provides effort while a non-restrictive OSS license over-provides effort, i.e., $e_r^* < e_o^* < e_{nr}^*$.

Proposition 2 supports the empirical finding that developer participation is lower under restrictive licenses than under non-restrictive licenses. However, the proposition highlights the fact that neither license provides optimal level of effort relative to cost. Public good characteristics of a restrictive license tend to depress effort below the efficient level. Here, developers, incentivized as users to participate, try to free-ride on each other's effort. On the other hand, a non-restrictive license generates revenue incentives as developers exert effort in order to become the winning proprietor of the software platform. This creates a tournament among developers to become the proprietor leading to over-investment by the losing developers. Thus, it is unclear whether the non-restrictive license is more optimal than the restrictive license.

3.4. License Choice

Let us now consider how a platform chooses an OSS in order to attract developers to participate. Competition to attract developers to the platform means that the equilibrium license maximizes developer surplus. As explained above, both licenses result in inefficient effort. Ex ante, developer surplus is thus maximized in the license with the lower inefficiency. License choice then facilitates the second-best efficient outcome for the platform.

Developer surplus under a restrictive license, given $e_r^* = 1$ is

Equation 4

$$S_r(N) = N - \frac{1}{\beta},$$

while surplus under the non restrictive license, given e_{nr}^* from Equation 3 is

$$S_{nr}(N, M) = (M + N)e_{nr}^* - \frac{1}{\beta} {e_{nr}^*}^{\beta},$$

which is

Equation 5

$$S_{nr}(N, M) = (M + N)^{\frac{\beta}{\beta-1}} \left(N - 1 + \frac{1}{N}\right)^{\beta-1} \left[1 - \frac{1}{\beta}\left(N - 1 + \frac{1}{N}\right)\right].$$

Surplus under a restrictive license is independent of M, i.e., the number of users who are not developers. However, surplus under non-restrictive OSS license increases with M. The proposition below characterizes license choice by a platform owner as a function of M and N.

Proposition 3

There exists $\widehat{N} > 1$ such that if $N \geq \widehat{N}$, then the platform owner chooses a restrictive license for all M. If $N < \widehat{N}$, then there exists $\widehat{M}$ such that the platform owner chooses a non-restrictive license if and only if $M > \widehat{M}$. The threshold, $\widehat{M}$, increases as the number of developers, N, becomes larger.

Proposition 3 states that non-restrictive licenses are more optimal when N is low and M is high. A high M implies that the platform has a large base of users who are not developers. Since developers do not have a way of appropriating platform value from these users in a restrictive license, there is a significant under-provision of effort relative to social value of the platform in such a license. On the other hand, a non-restrictive license allows developers to care about every user's value since it generates revenues for the proprietor. Here, this under-provision is absent.

At the same time, as the number of developers, N, rises, the probability of being the critical developer decreases. This reduces the expected surplus for developers under a non-restrictive license. Thus, a restrictive license performs well when there is a large number of developers but few users. Moreover, as the number of developers increases, the surplus that a non-restrictive license generate becomes relatively smaller and smaller. As this happens, the threshold number of users required to make the non-restrictive license attractive to developers needs to be larger and larger. Proposition 3 also provides a testable hypothesis. If license choice is endogenous, we should see greater proliferation of non-restrictive licenses in platforms with a large user base and relatively few developers.

Although we have only considered two main project characteristics for our analysis - namely, number of developers and the consumer base for the end product, there are several other factors that would influence developer participation decision and ultimately license choice. For example, one could imagine that quality of documentation influences the cost of effort. Or, the ideological perspectives of user-developers may make certain licenses more appealing. Another factor that influences license choice is the licensing of other complementary products. The choice of license affects how other complementary software with a different license can be combined. For example, GPL licensed software is highly restrictive in allowing the combination of non-GPL licensed code. Thus, if much of the complementary software is licensed under such restrictive licenses, new software that may need to be used in conjunction with other open source applications may also have to be licensed under a restrictive license. Although we have not explicitly examined these other factors, our model is sufficiently flexible to accommodate them. Some of these factors, such as quality of documentation affect developer participation but are likely to be uncorrelated with the type of license; in other words, there is no reason to believe that restrictive licenses result in better or worse documentation than non-restrictive licenses. If that is the case, Proposition 3 will continue to hold *ceteris parabus* (holding documentation quality across projects the same.) On the other hand, it is reasonable to expect that ideological motivations of developers and licensing of complementary software do influence license choice. If developers are driven by Stallman's open software philosophy, the under-provision inefficiency in restrictive licenses will be mitigated, thus creating a more favorable case for such licenses. In our model, we will see this in the form of a lower $\widehat{N}$ cut-off and a higher $\widehat{M}$ threshold for choosing a non-restrictive license. A similar argument holds for a situation where complementary software is developed under a restrictive license. Once again, this will create a higher incentive to choose a restrictive license.

4. Conclusion and Future Analysis

We provided a preliminary model of developer participation under different OSS licenses to explain the observed empirical phenomenon of greater developer participation under non-restrictive licenses. We also showed that both types of licenses result in inefficient participation. A platform owner choosing a license for an OSS platform trades-off these inefficiencies across licenses and chooses a non-restrictive license when there are a small number of developers or a very large user base.

There are a number of ways in which the framework presented here can be extended to provide a more complete understanding of platform licensing. One productive line of inquiry is to understand how developer and licensing incentives are formed in a competitive platform market. Following Microsoft's antitrust battles during the early 2000s, much of the literature on platform competition has focused on demand-side concerns of network efforts creating entry barriers. Google is facing similar allegations currently from EU over the bundling of applications with its Android OS on its phones.[5] Two crucial differences in Google's case are that Google has a non-restrictive OSS license and, unlike Microsoft in the past, it faces competition from Apple's proprietary iOS platform. The framework we have developed here can be extended to consider supply-side externalities created by an OSS platform and the role that licensing plays in the formation of a platform market's competitive landscape.

APPENDIX

Proof of Proposition 1

Comparing $e_r^* = 1$ and $e_o^* = 1$ from Equation 1 we see that $e_r^* - e_o^* = 1 - (M+N)^{\frac{1}{\beta-1}}$. Since $(M+N) > 1$, it must be the case that $e_r^* - e_o^* < 0$.[6] ■

Proof of Proposition 2

We look for a symmetric equilibrium where all users exert identical effort. Under a non-restrictive license, suppose all developers, $j \neq i$, choose $e_j = e_{nr}$. Then, surplus to developer i depends on whether she is a critical developer. Hence, surplus is:

$$E\left[S_i\left(e_i; e_{j\neq i}\right)\right] = \Pr[i=k]\,(M+N)\left(e_i + \sum_{j\neq i} e_j\right) - \frac{1}{\beta}e_i^{\beta}.$$

Let us now derive $\Pr[i=k] = \Pr\,(y_i > y_j, for\ all\ j \neq i)$. Given $e_j = e_{nr}$ for all $j \neq i$, developer i makes the critical contribution if and only if $e_i\varepsilon_i > e_{nr}\varepsilon_j$ for every $j \neq i$, i.e., $\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i$. Given ε_i, $\Pr\left(\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i \forall j \neq i\right) = \left[\Pr\left(\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i\right)\right]^{N-1}$. Now consider the following cases:

a. If $e_i < e_{nr}$, then $\frac{e_i}{e_{nr}}\varepsilon_i \leq 1$, so that $\Pr\left(\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i\right) = \frac{e_i}{e_{nr}}\varepsilon_i$, and

$$\Pr(i=k) = \int_0^1 \left(\frac{e_i}{e_{nr}}\varepsilon_i\right)^{N-1} d\,\varepsilon_i = \frac{1}{N}\left(\frac{e_i}{e_{nr}}\right)^{N-1}.$$

b. If $e_i \geq e_{nr}$, then $\frac{e_i}{e_{nr}}\varepsilon_i \leq 1$ if and only if $\varepsilon_i \leq \frac{e_{nr}}{e_i}$. Thus, for $\varepsilon_i \leq \frac{e_{nr}}{e_i}$, $\Pr\left(\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i\right) = \frac{e_i}{e_{nr}}\varepsilon_i$ and for $\varepsilon_i > \frac{e_{nr}}{e_i}$, $\Pr\left(\varepsilon_j < \frac{e_i}{e_{nr}}\varepsilon_i\right) = 1$.

$$\Pr(i=k) = \int_0^{\frac{e_{nr}}{e_i}} \left(\frac{e_i}{e_{nr}}\varepsilon_i\right)^{N-1} d\,\varepsilon_i + \int_{\frac{e_{nr}}{e_i}}^1 d\,\varepsilon_i = \frac{1}{N}\left[N - (N-1)\left(\frac{e_{nr}}{e_i}\right)\right].$$

Thus, expected surplus to developer i is:

$$E[S_i(e_i; e_{nr})] = \begin{cases} \dfrac{(M+N)}{N}\left(\dfrac{e_i}{e_{nr}}\right)^{N-1}[e_i + (N-1)e_{nr}] - \dfrac{1}{\beta}e_i^{\beta}\ if\ e_i < e_{nr}, \\ \dfrac{(M+N)}{N}\left[Ne_i + (N-1)^2 e_{nr} - (N-1)^2\dfrac{e_{nr}^2}{e_i}\right] - \dfrac{1}{\beta}e_i^{\beta}\ if\ e_i \geq e_{nr}. \end{cases}$$

Differentiating with respect to e_i, we get

$$\frac{d}{de_i}E[S_i(e_i; e_{nr})] = \begin{cases} \dfrac{(M+N)}{N}\left(N\left(\dfrac{e_i}{e_{nr}}\right)^{N-1} + (N-1)^2\left(\dfrac{e_i}{e_{nr}}\right)^{N-2}\right) - e_i^{\beta-1}\ if\ e_i < e_{nr}, \\ \dfrac{(M+N)}{N}\left[N - (N-1)^2\left(\dfrac{e_{nr}}{e_i}\right)^2\right] - e_i^{\beta-1}\ if\ e_i \geq e_{nr}. \end{cases}$$

Note that if $e_{nr} = 0$, then $\frac{d}{de_i}E[S_i(e_i; e_{nr})] \to \infty$ so that $e_i^* > 0$. Hence, we know that any symmetric equilibrium must have positive effort level, i.e., $e_{nr} > 0$.

For $e_i \geq e_{nr}$, the surplus function is concave because $\frac{d^2}{de_i^2}E[S_i(e_i; e_{nr})] < 0$. At $e_i = e_{nr}$, $\frac{d}{de_i}E[S_i(e_i; e_{nr})] = \left[(M+N)\frac{N^2-N+1}{N} - e_{nr}^{\beta-1}\right]$. So if $\frac{d}{de_i}E[S_i(e_i; e_{nr})] < 0$, i.e., if $e_{nr} > \left[(M+N)\frac{N^2-N+1}{N}\right]^{\frac{1}{\beta-1}}$, then $e_i^* < e_{nr}$. But this cannot be a symmetric equilibrium. Thus, a necessary condition is that $\frac{d}{de_i}E[S_i(e_i; e_{nr})] > 0$ at $e_i = e_{nr}$, or $e_{nr} \leq \left[(M+N)\frac{N^2-N+1}{N}\right]^{\frac{1}{\beta-1}}$.

The first order condition yields $\frac{(M+N)}{N}\left[N - (N-1)^2\left(\frac{e_{nr}}{e_i^*}\right)^2\right] = (e_i^*)^{\beta-1}$. Setting $e_i^* = e_{nr}$, we get:

$$e_{nr}^* = \left[(M+N)\frac{N^2-N+1}{N}\right]^{\frac{1}{\beta-1}}.$$

Let us check if developer surplus is positive here. Developer surplus is:

$$S_{nr}^* = (M+N)e_{nr}^* - \frac{1}{\beta}(e_{nr}^*)^{\beta}.$$

$S_{nr} > 0$ if and only if $e_{nr}^* < [\beta(M+N)]^{\frac{1}{\beta-1}}$, i.e., if $\beta > \frac{N^2-N+1}{N}$. For low enough N, say for $N \leq \beta$ (and $\beta \geq 2$ because $N > 1$), this condition holds.

Comparing the effort levels across the licenses and with efficient effort, we can see that $e_r^* < e_o^* < e_{nr}^*$.■

<u>Proof of Proposition 3</u>

Looking at Equation 4 and Equation 5, we see that while S_r is always positive, S_{nr} becomes negative when N is large enough, because $\left[1 - \frac{1}{\beta}\left(\frac{N^2-N+1}{N}\right)\right]$ is decreasing in N and becomes negative for large N. When $N = 1$, it is positive. Hence, for every β, there exists $\hat{N} > 1$ such that $\left[1 - \frac{1}{\beta}\left(\frac{\hat{N}^2-\hat{N}+1}{\hat{N}}\right)\right] = 0$; hence for all $N \geq \hat{N}$, we have $S_{nr} \leq 0 < S_r$ and thus a restrictive license is chosen.

Now let us consider, for a given β, $N < \hat{N}$, hence $S_{nr} > 0$. Note that S_r does not depend on M while $\frac{dS_{nr}}{dM} > 0$. At $M = 0$ and $N = 1$, $S_{nr} = S_r$. Hence, for every N, there must exist a large enough M, say $\hat{M}$, so that $S_{nr} > S_r$ for all $M > \hat{M}$ and a non-restrictive license is chosen. Since S_{nr} is decreasing in N, $\hat{M}$ must increase with N for all $N < \hat{N}$.■

ENDNOTES

1. See Sen et al. (2008), Comino et al. (2007), Lerner and Tirole (2005), Fershtman and Gandal (2007).
2. Open Source Software development often involves simultaneous production and consumption of the software since programmers who work on the code often also use the modifications they make on the software. Hence the term “user-developer” is used when modelling OSS development. See Atal and Shankar (2014), Athey and Ellison (2014) and Johnson (2002).

3. See Appendix, Proof of Proposition 2, for derivation of this probability density function.

4. See Proof of Proposition 2 in the appendix for details.

5. https://arstechnica.com/gadgets/2022/09/eu-upholds-googles-e4-1-billion-fine-for-bundling-search-with-android/

6. $e_r^* = e_o^*$ when $M = 0$ and $N = 1$, but this is not an interesting case.

REFERENCES

Atal, V. and K. Shankar (2015) "Developers' Incentives and Open Source Software Licensing: GPL vs. BSD," *B.E. Journal of Economic Analysis and Policy*, Vol. 15, No. 3: 1381-1416.

Atal, V. and K. Shankar (2014) "Open Source Software: Competition with a Public Good," *Atlantic Economic Journal*, Vol. 42, No. 3: 333-345.

Athey, S. and G. Ellison (2014) "Dynamics of Open Source Movements," *Journal of Economics and Management Strategy*, Vol. 23, No. 2: 294-316.

Comino, S., F. M. Manenti and M. L. Parisi (2007) "From planning to mature: On the success of open source projects," *Research Policy*, Vol. 36, No. 10: 1575-1586.

Fershtman, C. and N. Gandal (2007) "Open Source Software: Motivation and Restrictive Licensing,"*International Economics and Economic Policy*, Vol. 4, No. 2: 209-225.

Hann, I.-H., J. A. Roberts, and S. A. Slaughter (2013) "All Are Not Equal: An Examination of the Economic Returns to Different Forms of Participation in Open Source Software Communities," *Information Systems Research*, Vol. 24, No. 3: 520--538.

Johnson, J. P. (2002) "Open Source Software: Private Provision of a Public Good," *Journal of Economics and Management Strategy*, Vol. 11, No. 4: 637-662.

Lakhani K., and R. Wolf (2005) "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects," In: Feller J., B. FitzGerald, S. Hissam, K. Lakhani, *Perspectives on Free and Open Source Software*, Cambridge: MIT Press.

Lerner, J. and J. Tirole (2005) "The Scope of Open Source Licensing," *Journal of Law, Economics and Organization*, Vol. 21, No.1: 20-56.

Lerner, J., and J. Tirole (2002) "Some Simple Economics of Open Source," *The Journal of Industrial Economics*, Vol. 50, No. 2: 197-234.

Roberts, J. A., I. -H. Hann, and S. A. Slaughter (2006) "Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects," *Management Science*, Vol. 52, No. 7: 984-999.

Sen, R., C. Subramaniam, and M. L. Nelson (2008) "Determinants of the Choice of Open Source Software License," *Journal of Management Information Systems*, Vol. 25, No. 3: 207-239.

Performance, Pandemic and Probable Peril: Disentangling Risk, Volatility and Bubbles in Faang Portfolio

Chitrakalpa Sen* and Gagari Chakrabarti†

ABSTRACT

In this study, we examine the risk characteristics of the FAANG stocks by constructing a minimum variance portfolio (MVFAANG). Google and Apple are found to have the highest portfolio weights. These FAANG stocks showed a remarkable growth in value till 2022, which increased even more rapidly during the pandemic. However, since 2022 these stocks experienced a precipitous decline in value. We further explore into the risk dynamics of MVFAANG by decomposing its total volatility in permanent and transitory components. Our results show that the portfolio's volatility entirely consists of a considerably high permanent volatility. Our findings also indicate that the value-at-risk of MVFAANG is also substantially high, reaching 25% during the pandemic. Finally, using GSADF model we identify multiple structural breaks and explosiveness in MVFAANG. Overall results indicate that MVFAANG possesses substantial risk, which calls for cautious consideration for the investors.

1. INTRODUCTION

Facebook (Meta Platforms Inc.), Amazon.com Inc, Apple Inc., Netflix Inc. and Google (Alphabet Inc), commonly known as the FAANG[i] stocks are the most well-known technology giants in the world.

The FAANG quintet, immensely popular with the investors, make up almost 16% of the entire value of S&P500 in 2023 (Meryn, 2023). The high prices of these stocks can be justified by the fact that these companies as innovation leaders in their respective industries have been instrumental in revolutionizing the technology landscape. These stocks experienced massive surge in growth in the recent past, consistently outperforming the market substantially (Hobbs and Suarez, 2022), and therefore can be classified as growth stocks. Growth stocks grow at a faster rate than the overall market, with a higher price-earning ratio than value stocks, although literature provides evidence that in the longer run, value stocks outperform growth stocks (Mukherji et. al., 1997; Fama and French, 1998; Chan and Lakonishok, 2004; Wang, 2011). However, growth stocks are found to be less risky than value stocks during bad times, when the expected market risk premium is high (Petkova & Zhang, 2003).

2. RESEARCH QUESTIONS

The price dynamics of the individual FAANG stocks reveals a rather paradoxical pattern. These stocks performed exceptionally well during the pandemic-induced crisis, demonstrating resilience and growth, but

* Jindal Global Business School, O. P. Jindal Global University, Sonipat-Narela Road, Sonipat, Haryana-131001, India, Email: csen@jgu.edu.in

† Department of Economics, Presidency University, 86/1 College Street, Kolkata 700073, India, Email: gagari.econ@presiuniv.ac.in

encountered a dramatic downturn when the economy began to recover. This behaviour of the FAANG stocks raises several important questions: What underlying factors caused this counterintuitive price behaviour? Are these growth stocks merely fad stocks or glamour stocks? What are the implications for investors and fund managers? To address these questions, and have greater insights into the paradox, we propose the following research questions.

Q1: How should the FAANG stocks be combined to construct a minimum-variance portfolio to satisfy the risk-averse investors?

Q2: Is the constructed minimum variance FAANG portfolio (MVFAANG) intrinsically risky? This question is further is explored in three ways:

Q2a. Do the permanent and the transitory risks of MVFAANG differ? While high permanent volatility is undesirable, significant transitory (or, short-term) volatility can lead to market instability.

Q2b. What is the potential loss from investing in such a portfolio?

Q2c. Are there multiple breaks and bubbles in the prices of MVFAANG?

In this paper, these research questions aim to shed light on the performance of MVFAANG, with an emphasis on the recent pandemic time period.

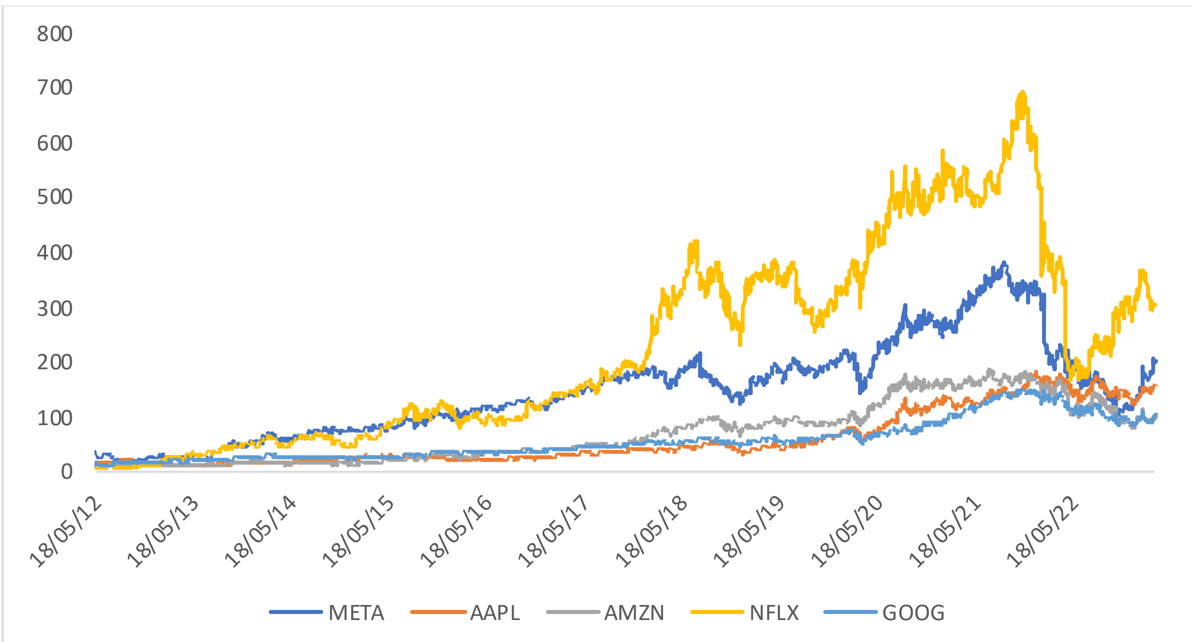
3. DATA AND BACKGROUND

For this study, we have used the daily prices for FAANG stocks from May 18, 2012, through March 21, 2023, providing a sample size of 2727 observations. The data has been collected from the Bloomberg database. For further in-depth analysis, we have considered the daily return of these stocks, calculated using the natural logarithm of the ratio of the stock price on day t (P_t) to the stock price on the preceding day (P_{t-1}). The daily return is expressed $R_i = ln(P_t/P_{t-1})$.

The selection of this sample period is supported by four main strategic considerations. First, by 2012 the US economy was well into recovery following the financial crisis of 2007-08. Second, Facebook (now Meta) went public with its initial public offerings on May 18, 2012. Extending our sample period till that date allows us to carry out a comprehensive analysis of the entire FAANG portfolio. Third, the 130 month-long sample period allows us to examine the gradual evolution in the FAANG stock prices, including a major event like the recent Covid-19 pandemic induced crisis. And finally, by extending the sample period through March 2023 allows us to examine the FAANG stock prices in a state of relative normalcy post the pandemic.

Figure-1 illustrates the movement of individual FAANG stock prices over the time horizon. It is evident that since the beginning of 2018 these stocks experienced a significant surge in prices. This increase has been more pronounced between early 2020 and late 2021. Among the five constituent stocks, Netflix and Meta have clearly led in performance (Figure-1). These stocks not only demonstrated remarkable resilience during the pandemic, they thrived (Sozzi, 2020). This stellar performance can be attributed to various factors, such as the digital acceleration as reliance on technology increased for remote work, increased social media engagement, increased consumption of home entertainment at home, or a surge in online shopping. However, in 2022, all five stocks suffered a substantial decline in the prices, with Netflix and Meta again leading the pack, with a 64% and 51% drop in price, respectively, while the other three stocks fell by at least 27%, contributing to a 19% fall in the S&P500, the worst since 2008 (Miao, 2023). Consequently, Apple posted a revenue of $117.2bn, down 5% year-on-year in fiscal 2023 first quarter (Apple, 2023), while Netflix, Meta, Alphabet (Google's parent company) and Amazon initiated a large scale layoff (Hern, 2023; Bushard, 2023; Turner, 2023). The steep climb in the price of the FAANG stocks leading up to 2022 followed by a rapid decline hints towards the possibility of these stocks being overvalued and characterized by bubble. A recovery appears to be underway, as the stock prices began to exhibit upward trends during the last months of 2022 and the beginning of 2023, with Netflix showing the earliest sign of recovery, as early as in July 2022.

Figure-1: FAANG[ii] Stocks – Daily price from 2012 – 2023



This apparently perplexing behaviour of the FAANG stocks therefore requires a close examination. In the next section, we discuss the methodology used to address the research questions.

4. METHODOLOGY:

4.1. CONSTRUCTION OF MINIMUM VARIANCE (MV) PORTFOLIO

We start our analysis by constructing a minimum variance portfolio with all five constituent stocks. The minimum variance portfolio with the FAANG stocks would provide valuable insights into these stocks' performance, price behaviour, underlying risk dynamics and potential diversification benefit. A risk-averse investor will seek out portfolios that minimise the risk given the expected return. This automatically makes the minimum variance portfolio the optimal portfolio for the risk-averse investor. The early pioneering research of Haugen and Baker (1991) found that investing in a stock portfolio that exposed the investors to minimum risk (variance) exhibited superior performance compared to the market-cap weighted index. The superior performance of minimum variance portfolio is well-documented and advocated in empirical literature (Winston, 1993; Chan et al. 1999; Ledoit and Wolf, 2003; Jagannathan and Ma, 2003; Kempf and Memmel, 2006; Clarke et. al., 2006, 2011; Frahm, 2008; Nielsen and Aylursubramanian, 2008; Scherer, 2010).

Assuming that (a) investors choose a portfolio to maximize a derived concave utility function of the form $V(\overline{Z},\sigma^2)$ with $V_2 < 0$ and $V_1 > 0$; (b) investors have a common time horizon and homogeneous belief about $\overline{z}$, and □ □□(c) each asset is infinitely divisible and (d) there is a riskless asset that could be bought or sold without any restriction, each investor will hold a minimum variance (MV) portfolio. No other portfolio could be optimal because given $V_2 < 0$, the minimum variance portfolio with the same expected return would be preferred to any other alternative. This addresses our research question Q1.

Given the return on the *i*'th asset (R_i), its standard deviation (σ_i) and the proportion of wealth invested in the *i*'th asset (X_i), portfolio return (R_{PF}) and risk (σ^2_{PF}) are given as:

$$R_{PF} = X'R = (X_1, X_2, \ldots, X_n) \cdot \begin{pmatrix} R_1 \\ R_2 \\ \ldots \\ R_n \end{pmatrix}$$

$\sigma^2_{(PF)} = var(X'R) = X'\Sigma X$ where Σ is the variance-covariance matrix of returns.

The choice of MV portfolio is essentially a choice of weights given to available assets. The MV portfolio $X = (X_1, X_2, \ldots, X_n)'$ in an n-asset case solves the constrained minimization problem:

$$\min_{X_1, X_2, \ldots X_n} \sigma^2_{PF} = X_1^2\sigma_1^2 + \cdots + X_n^2\sigma_n^2 + 2\sum_{i,j} X_i X_j \sigma_{ij} = X'\Sigma X$$

$$\text{Such that } X_1 + X_2 + \ldots + X_n = 1 \text{ or, } X'1 = 1$$

4.2. DECOMPOSING TOTAL VOLATILITY INTO PERMANENT AND TRANSITORY COMPONENT: COMPONENT GARCH MODEL

The conditional variance in a GARCH (1,1) model shows volatility mean-reversion to a constant. By contrast, the CGARCH model allows mean-reversion to a varying level. A CGARCH model is used to decompose the total volatility into a component capturing the long term or the permanent changes in volatility; and a component capturing the short term or the transitory fluctuations in volatility (Engle and Lee, 1999). The permanent component represents the underlying systematic non-diversifiable risk, while the transitory component represents the unsystematic, diversifiable risk. The decomposition of total risk into these two components provides greater insights into the risk dynamics of the portfolio. If the majority of the total risk is attributed to the permanent component, it implies that the risk is inherent in nature and cannot be diversified away. Investors need to be cautious and they need to take into consideration broader market aspects while assessing the risk and investing. On the other hand, a high transitory risk indicates the necessity of more efficient diversification of the investment.

In the CGARCH model, the conditional volatility of daily returns of MVFAANG is -

$\sigma_t^2 = \overline{\omega} + \alpha(\varepsilon_{t-1}^2 - \overline{\omega}) + \beta(\sigma_{t-1}^2 - \overline{\omega})$ which is decomposed into

$\sigma_t^2 - m_t = \alpha(\epsilon_{t-1}^2 - m_{t-1}) + \beta(\sigma_{t-1}^2 - m_{t-1})$ - Transitory component

and

$m_t = \omega + \rho(m_{t-1} - \omega) + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$ - Permanent component

While volatility reverts to a constant $\overline{\omega}$ in the GARCH (1,1) model, CGARCH allows mean-reversion to a long-run time-varying volatility m_t. The transitory, or the short-run component $(\sigma_t^2 - m_t)$ converges to zero with powers of ($a + b$). The long-run component m_t converges to ω with powers of r.

This addresses our research question Q2a.

4.3. DYNAMIC VALUE-AT-RISK

Next, we use the Value-at-risk (VaR) method to examine the maximum loss to be incurred with a certain probability. Dynamic VaR, that incorporates stylised facts of financial time-series, is calculated as:

$$\mathrm{VaR(asset)} = \mathrm{conditional\ mean} + \mathrm{conditional\ variance} * \mathrm{F}^{-1}(\mathrm{a})$$

$F^{-1}(a)$ is the inverse pdf of t-distribution and conditional mean and conditional variance are estimated from a suitable GARCH model. This section of the methodology address the research question Q2b.

4.4. TESTING FOR STRUCTURAL BREAKS AND EXPLOSIVENESS

A comprehensive examination of potential breaks and bubbles in MVFAANG can provide an in-depth understanding of the intrinsic risk dynamics. The presence of structural breaks is one of the stylized facts in financial market. A structural break is said to have occurred if at least one of the parameters of a time series makes paradigm shift at some point that is permanent and irreversible (Andrews, 1993; Andrews and Ploberger, 1994; Bai and Perron, 1998, 2003, Hansen, 2001; Brooks, 2002).

We test multiple endogenous structural breaks in the daily price series of MVFAANG. To do this, we use the Bai-Perron multiple breakpoint test method (2003. This method uses the M-fluctuation test, known for its robustness in testing parameter instability (Zeileis and Hornik, 2007).

An important form of nonstationarity in time series is the presence of explosive roots. They indicate periods of rapid and significant increase in an asset price, often diverging from its fundamental value over time (Phillips et al., 2011; Phillips et al. 2015; Ashtill et. al., 2017, 2018). Periods of explosive behaviour are commonly associated with speculative bubbles (Diba and Grossman, 1988). Speculative bubbles arise in the presence of exuberant market behaviour, when investors disregard the fundamentals expecting that prices will persistently increase and profits can be made in future (Santoni, 1987). This causes a divergence between market prices and the fundamental values. Positive bubble prices will further increase expectations, causing greater divergence from the fundamental values. Increase in this divergence causes the intrinsic risk of the asset to increase fast, before the bubble finally bursts. And when the bubble bursts, investors holding the asset suffer significant losses (Anderson and Brooks, 2012). Presence of speculative bubbles in stock market has been documented by an extensive body of literature (Anderson and Brooks, 2010, Brooks and Katsaris, 2003, 2005; Costa et al. 2017).

To identify any explosive behaviour in MVFAANG prices, which may lead to potential bubbles, the GSADF model is used. Phillips et al. (2011) developed the right-tailed versions of the ordinary augmented-Dickey–Fuller test with the parameter d to test the null hypothesis that d = 1 (so that the time series contains a unit root) against the alternative hypothesis of d > 1 (so that time series contains an explosive root). Explosive behaviour exists if the root (d) exceeds 1 but approaches 1 as the sample size increases (Phillips and Magdalinos, 2007a, 2007b). The test includes a recursive supremum ADF (SADF) statistic, and a generalized supremum ADF (GSADF) test statistic as follows:

$$SADF(r_0) = sup_{r_2 \in [r_0,1]} \{ADF_0^{r_2}\}$$

$$GSADF(r_0) = sup_{r_2 \in [r_0,1], r_1 \in [0, r_2 - r_0]} \{ADF_{r_1}^{r_2}\}$$

r_1, r_2 ∈ [$0,1$] are a series of subsample. A rolling window ADF regression is used following Corbet et al. (2018) and Bouri et al. (2019). We start from fraction r_1 and end at r_2. With r_w as the rolling window, $r_2 = r_1 + r_w$. This section of the methodology addresses the research question Q2c.

5. RESULT AND DISCUSSION

Google and Apple dominate the MV portfolio. MVFAANG assigns weights of 0.46 to Google, 0.37 to Apple, 0.13 to Amazon, 0.03 to Netflix and 0.01 to Meta. Portfolio value has increased significantly over time. The fall during the pandemic was not significant and the value increased sharply since March 2020. Prices fell in 2022, but recovery started in January, 2023.

Figure-2: Movement in FAANG portfolio value

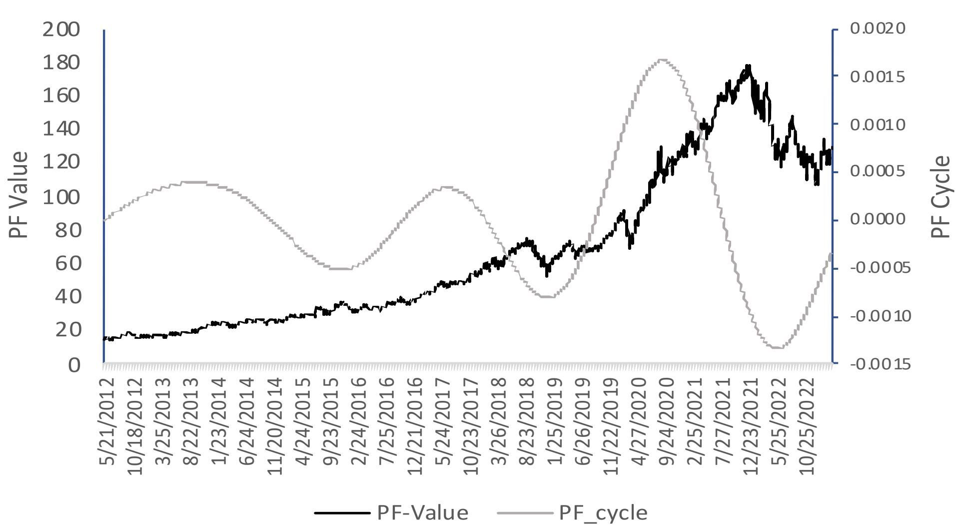
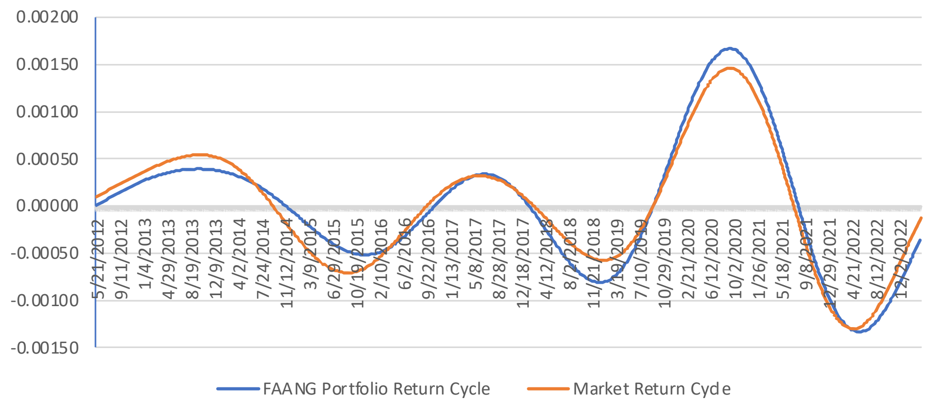


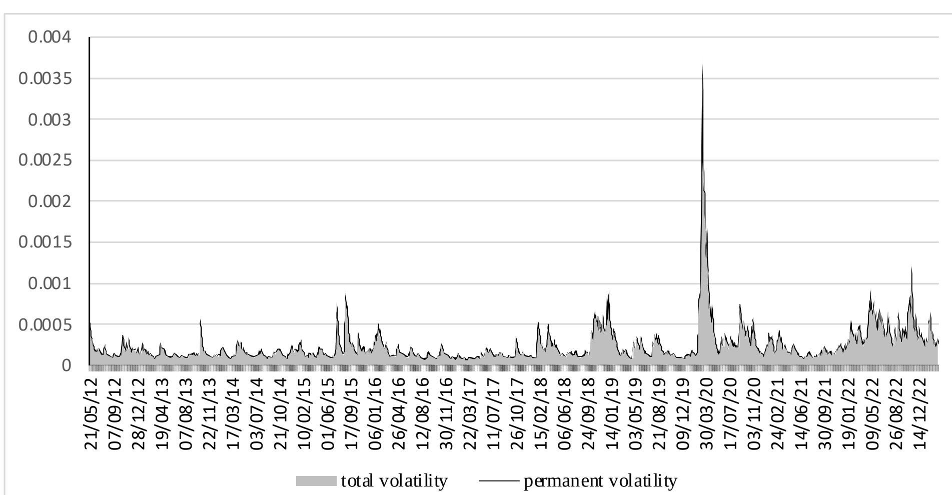
Figure-3 shows the cycles of return from the FAANG portfolio and the market (NASDAQ composite). During the period from 2012 to 2016, the FAANG portfolio acted as a hedge. When the market was falling since 2013, FAANG portfolio was giving more return than the market. This protection against the downside risk came at the expense of a ceiling on the FAANG portfolio return in a bullish market. Its return was lower than the market during 2012-2013 – a behaviour that is typical to a hedging instrument. The nature of the portfolio however has changed over time. From 2016 to 2018, its returns closely resembled that of the market. From early-2019 to early-2022 FAANG portfolio dominated the market in bullish as well as in bearish condition. In recent years, however, the market has started dominating the portfolio.

Figure-3: MVFAANG return cycle vs the market return cycle



Total volatility of portfolio return is decomposed into permanent and transitory components (Figure-4). The vertical axis shows the magnitude of total volatility calculated from the CGARCH model. The black line shows the magnitude of the permanent volatility. The difference between these two is the transitory volatility. We can see from the figure that almost all of the total volatility comes from the permanent volatility, and the transitory component is negligible. This points towards the fact that the risk in the portfolio is systematic and non-diversifiable.

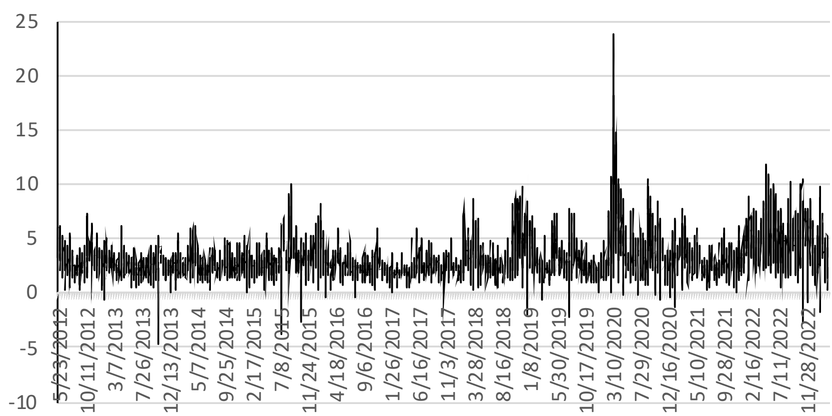
Figure-4: Permanent and Transitory components of FAANG portfolio volatility



Portfolio risk was low most of the times, except during pandemic. Although the transitory risk was never significant, it does not help risk-avert investors as the permanent volatility is quite high during pandemic.

Similarly, the VaR was significantly high during pandemic, touching almost 25%. Although the VaR or the risk of loss was quite low in other periods, this sudden increase in loss during pandemic is not desirable. This adds to the risk of investment (Figure-5).

Figure-5: Dynamic VaR for FAANG portfolio



Bai-Perron breakpoint test identifies multiple structural breaks in the MVFAANG series. The M-fluctuation test statistic is 21.631 (Table-1), significant at 1% level, providing strong evidence to reject the null hypothesis of absence of structural breaks.

Table - 1: Bai-Perron M-Fluctuation Test for Structural Breaks in MVFAANG

Test	Test Statistic	P-value
M – fluctuation	21.631	0.000

Further, the SADF and GSADF tests detect presence of explosiveness and multiple breaks in the portfolio price. The null hypothesis is rejected (Table-2) and presence of explosive behavior is confirmed. This could be an indication of the existence of a speculative bubble.

Table-2: SADF and GSADF test statistic

		Critical Values		
Test Statistic		**90%**	**95%**	**99%**
GSADF	3.49	-15.08	-14.69	-14.12
SADF	2.46	-50.84	-50.51	-49.96

6. CONCLUSION

In order to conduct an in-depth analysis of the FAANG stocks' inherent risk, we constructed a minimum variance portfolio (MVFAANG), which is dominated by Google and Apple. During the pandemic, the value of this portfolio spiked sharply, before plummeting in 2022. We further decomposed the total volatility into permanent and transitory components and found that although volatility was quite low during most of the period, it rose significantly during the pandemic. Moreover, total volatility is found to be driven almost entirely by a permanent volatility of relatively high magnitude. The results further indicate that the value-at-risk to be significantly high during pandemic, nearly at 25%, but it remained low during other periods. This sudden increase in loss during pandemic is undesirable. Furthermore, our analysis found the presence of multiple structural breaks and explosiveness in the portfolio price. This indicates a higher degree of uncertainty and risk in the portfolio.

The FAANG stocks are recognised as growth stocks that enjoyed significant escalation in price in the pandemic and post-pandemic periods. Historically growth stocks (like IT stocks and more recently, the green stocks) had often enjoyed such escalation that was justified by the four most dangerous words in finance – "this time it's different". Higher valuations of growth stocks tend to be justified, not in terms of strong fundamentals, but by their perceived growth opportunity. Non-realization of such perception has often led to panic and crashes. Our study has spotted the presence of an explosive behaviour in the FAANG portfolio prices. The pandemic era has undoubtedly opened up growth opportunities for these stocks, but a presence of speculative bubble cannot be denied. The presence of high permanent volatility and loss from investment during pandemic escalated the concern.

This paper makes several unique contributions to the existing literature. First, the risk dynamics of FAANG stocks, especially during and immediately after the COVID-19 crisis, has received little scholarly attention. This paper tries to fill that gap. Second, use of component GARCH model allows a more granular exploration of the portfolio's risk profile. By decomposing the total volatility into permanent and transitory components, we provide a novel perspective into the risk dynamics of the FAANG stocks. Lastly, the findings have significant implications for investors and fund managers with exposure in FAANG stocks. Such characteristics indicate that the inherent risk in these stocks might be more substantial from what may be apparent from a cursory analysis of historical returns. Therefore, investors and fund managers should exercise caution while investing in these stocks, with particular attention to Apple and Google, given their dominant influence on the portfolio's overall risk. Periods of uncertainty or crisis call for even higher caution, with a need to diversify across sectors and implementing robust risk management strategies to mitigate the potential high risk in these stocks.

ENDNOTES

1. The acronym, FANG (without Apple) was first used by the TV Host and founder of TheStreet Jim Cramer in CNBC's "Mad Money". Apple was a late entrant in 2017, making the acronym FAANG.
2. The NASDAQ ticker symbols for the companies are - Facebook or META : META, Amazon.com Inc: AMZN, Apple Inc.: AAPL, Netflix Inc.: NFLX, Google or Alphabet Inc: GOOG.

REFERENCES

Anderson, K., Brooks, C. and Katsaris, A. (2010). *Speculative bubbles in the S&P 500: was the tech bubble confined to the tech sector?* Journal of Empirical Finance, 17 (3). pp. 345-361.

Anderson, K. A., & Brooks, C. (2012). *Speculative Bubbles and The Cross-sectional Variation In Stock Returns*. SSRN Journal. https://doi.org/10.2139/ssrn.2176444

Andrews, D.W.K. (1993). *Tests for Parameter Instability and Structural Change with Unknown Change Point*. Econometrica, 61, 821–856.

Andrews, D.W.K., Ploberger, W. (1994). *Optimal tests when a nuisance parameter is present only under the alternative*, Econometrica 62, 1383–1414.

Apple. (2023, March 23). *Apple reports first quarter results*. Apple Newsroom. Retrieved March 27, 2023, from https://www.apple.com/newsroom/2023/02/apple-reports-first-quarter-results/

Bai J., & Perron, P. (1998). *Estimating and Testing Linear Models with Multiple Structural Changes*, Econometrica, Econometric Society, 66(1), 47-78, January.

Bai J., & Perron, P. (2003). *Computation and Analysis of Multiple Structural Change Models*, Journal of Applied Econometrics, 18(1), 1-22.

Bouri, E., Shahzad, S. J. H., & Roubaud, D. (2019). *Co-explosivity in the cryptocurrency market*, Finance Research Letters, 29, 216-221.

Brooks, C. (2002). *Introductory Econometrics for Finance*, Cambridge University Press.

Brooks, C., & Katsaris, A. (2003). *Rational Speculative Bubbles: An Empirical Investigation of the London Stock Exchange*. Bulletin of Economic Research, 55(4), 319–346. doi:10.1111/1467-8586.00179

Brooks, C. and A. Katsaris (2005), *A three-regime model of speculative behaviour: modelling the evolution of bubbles in the S&P 500 Composite Index*, Economic Journal, 115 (505), 767- 797.

Bushard, B. (2023, March 29). *2023 Layoff Tracker: Electronic Arts, Warner Music And Automaker Lucid Cut Jobs*. Forbes. Retrieved March 30, 2023, from https://www.forbes.com/sites/brianbushard/2023/03/29/2023-layoff-tracker-warner-music-cuts-270-jobs-while-electric-car-company-lucid-slashes-1300/?sh=7d1cd057115e

Chan, L.K.C., J. Karceski, and J. Lakonishok (1999). *On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model*, Review of Financial Studies, 12, 937-974.

Chan, L. K. C., & Lakonishok, J. (2004). *Value and growth investing: Review and update*. Financial Analysts

Journal, 60(1), 71-86.

Clarke, R., de Silva, H., Thorley, S. (2006). *Minimum-variance portfolios in the U.S. equity market*. Journal of Portfolio Management. 33(1), 10–12.

Clarke, R., de Silva, H., Thorley, S. (2011). *Minimum-variance portfolio composition*. Journal of Portfolio Management. 37(2), 31–45.

Corbet, S., Lucey, B., & Yarovaya, L. (2018). *Datestamping the Bitcoin and Ethereum Bubbles*, Finance Research Letters 26: 81-88.

Costa, C. T., da Silva, W. V., de Almeida, L. B., da Veiga, C. P. (2017). *Empirical evidence of the existence of speculative bubbles in the prices of stocks traded on the São Paulo Stock Exchange*. Contaduría y Administración, 62(4), 1317-1334. https://doi.org/10.1016/j.cya.2017.02.007

Engle, R. & Lee, G. (1999). A permanent and transitory component model of stock return volatility. In Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger; Engle, R., White, H., Eds.; Oxford University Press: Oxford, UK, 1999; 475–497.

Fama, E. F., & French, K. R. (1998). *Value versus Growth: The International Evidence*. The Journal of Finance, 53(6), 1975–1999. http://www.jstor.org/stable/117458

Frahm, G. (2008). *Linear statistical inference for global and local minimum variance portfolios*. Statistical Papers, DOI: 10.1007/s00362-008-0170-z

Haugen, R., Baker, N. (1991). *The efficient market inefficiency of capitalization-weighted stock portfolios*. Journal of Portfolio Management, 17, 35–40.

Hansen, B. E. (2001) *The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity*, Journal of Economic Perspectives, 15(4), 117–128.

Hern, A. (2023, March 14). *Zuckerberg's Meta to lay off another 10,000 employees*. The Guardian. Retrieved March 24, 2023, from https://www.theguardian.com/technology/2023/mar/14/mark-zuckerberg-meta-layoffs-hiring-freeze

Hobbs, B., & Suarez, J. (2022, July 26). *FAANG: An acronym that stands for five very successful tech companies that can move the stock market*. Business Insider. Retrieved March 14, 2023, from https://www.businessinsider.com/personal-finance/what-is-faang?IR=T

Jagannathan, R., Ma, T. (2003). *Risk reduction in large portfolios: why imposing the wrong constraints helps*, Journal of Finance 58, 1651–1683.

Kempf, A., Memmel, C. (2006). *Estimating the global minimum variance portfolio*, Schmalenbach Business Review 58, 332–348.

Ledoit, O., Wolf, M. (2003). *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection*, Journal of Empirical Finance 10, 603–621.

Meryn, R. (2023, March 20). *From Silicon Valley To Wall Street: FAANG Stocks Ruled The Market*. Industry Leaders Magazine. Retrieved March 24, 2023, from https://www.industryleadersmagazine.com/from-silicon-valley-to-wall-street-faang-stocks-ruled-the-market/

Miao, H. (2023, January 1). *The Year Big Tech Stocks Fell From Glory*. The Wall Street Journal. Retrieved

March 23, 2023, from https://www.wsj.com/articles/the-year-big-tech-stocks-fell-from-glory-11672543952

Mukherji, S., Dhatt, M.S., & Kim, Y.H. (1997). *A Fundamental Analysis of Korean Stock Returns*. Financial Analysts Journal, 53, 75-80.

Nielsen, F., & Aylusubramanian, R. (2008). *Far from the madding crowd – volatility efficient indices*. MSCI Barra Research Insights.

Petkova, R., & Zhang, L. (2003). *Is value riskier than growth*? NBER, Retrieved March 23, 2023, from https://users.nber.org/~confer/2003/si2003/papers/ap/zhang.pdf

Phillips, P. C. B., Wu, Y., & Yu, J. (2011). *Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?*, International Economic Review 52: 201-226.

Phillips, P.C.B., Magdalinos, T. (2007a). Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136, 115–130. https://doi.org/10.1016/j.jeconom.2005.08.002

Phillips, P.C.B., Magdalinos, T. (2007b). The Refinement of econometric estimation and test procedures: Finite sample and asymptotic analysis, Chapter: *Limit theory for moderate deviations from a unit root under weak dependence*, 123–162. Cambridge University Press, Cambridge. https://doi.org/10.1017/cbo9780511493157.008

Santoni G. J. (1987). *The Great Bull Markets 1924-29 and 1982-87: Speculative Bubbles or Economic Fundamentals?*, Federal Reserve Bank of St. Louis, Nov., 16-29.

Scherer, B. (2010). A New Look at Minimum Variance Investing. Social Science Research Network. [Retrieved 2014-12-15] Available at: http://ssrn.com/abstract=1681306

Sozzi, B. (2020, May 19). *FAANG stocks are totally ignoring the COVID-19 pandemic and are approaching this stunning level*. Yahoo! Finance. Retrieved March 15, 2023, from https://finance.yahoo.com/news/faang-stocks-are-totally-ignoring-the-covid-19-pandemic-and-are-approaching-this-stunning-level-173100962.html

Turner, J. (2023, March 23). *Tech Companies That Have Made Layoffs in 2023*. Tech.co. Retrieved March 30, 2023, from https://tech.co/news/tech-companies-layoffs

Wang, J. (2011). *Impact of investment horizon on the performance of value versus growth styles and style allocation*. Journal of Asset Management, 12, 438-446.

Winston, K. (1993). *The efficient index and prediction of portfolio variance*. Journal of Portfolio Management, 19, 27–34.

Zeileis, A., & Hornik, K. (2007). *Generalized M-fluctuation tests for parameter instability*. Statistica Neerlandica, 61(4), 488–508. doi:10.1111/j.1467-9574.2007.00371.x

Block Share Purchases and Firm Performance

Eleni Mariola[*], Katarzyna Platt[†], Elena Smirnova[‡], Frank Sanacory[§]

ABSTRACT

Blockholders acquire a significant interest in firms to influence management to take actions to improve firm value. We investigate the impact of block share purchases on firm value and performance. Using a propensity score-matched sample, we find that firms targeted by individuals, partnerships, investment advisors, and corporations experience deteriorating operating performance for up to two years after the intervention. The results also indicate that firms targeted by multiple blockholders in the same year experience a smaller decrease in firm performance than those subject to a single campaign. We find no evidence that blockholders campaigns lead to improved long-term performance.

INTRODUCTION

For several decades, blockholders and activist investors have played a significant role in corporate restructuring, takeovers, and acquisitions. Shareholder activism occurs when certain stockholders, dissatisfied with how management runs the company, leverage their rights as owners to alter the company's strategy. Corporate governance, executive compensation, shareholder value, say on pay, and environmental and social proposals include some of the issues addressed by activist investors.

The effectiveness of blockholding and activism has been evaluated by the prior literature (for a detailed analysis of prior empirical studies refer to Gillan and Starks (2009) and Denes et al. (2017)). Previous studies examine the market reaction to the activist campaign and the effect on the target's operating performance, both short-term (days around the intervention) and the long-term (up to 5 years after intervention). Zenner et al. (2005); Klein and Zur (2009); Brav et al. (2008) find short-term positive abnormal returns following an activist campaign. Bebchuk et al. (2015) and deHaan et al. (2019) extend the post intervention study period to five years and report positive abnormal returns. Cremers et al. (2020) show that both target and control firms experience positive abnormal returns. Some researchers showed statistically insignificant results (Smith, 1996; Wahal, 1996; deHaan et al., 2019), others found no improvement in operating performance (Gillan, 1995; Karpoff et al., 1996; Del Guercio and Hawkins, 1999; Klein and Zur, 2009), while a third group of researchers found increased operating performance following an intervention

[*] Department of Finance, Business Economics and Legal Studies, LaPenta School of Business, Iona University, 715 North Avenue, New Rochelle, NY 10801. Emariola@iona.edu

[†] Department of Management, Marketing and Finance, School of Business, State University of New York at Old Westbury, P.O. Box 210, Old Westbury, NY 11568. PlattK@oldwestbury.edu

[‡] Department of Management, Marketing and Finance, School of Business, State University of New York at Old Westbury, P.O. Box 210, Old Westbury, NY 11568. SmirnovaE@oldwestbury.edu

[§] Department of Mathematics, Computer and Information Science, School of Arts and Sciences, State University of New York at Old Westbury, P.O. Box 210, Old Westbury, NY 11568. SanacoryF@oldwestbury.edu

(Brav et al., 2008; Bebchuk et al., 2015).

Our study contributes to the literature on blockholders and corporate performance in several ways. First, we extend prior studies by investigating the effects of blockholding employed by different types of investors on corporate performance. Second, we study whether the purpose for the block share purchase, as disclosed in Item 4 of Schedule 13D, impacts the target's future performance. We investigate whether the interest activists acquire in target firms enables them to significantly influence firm management, including operational and strategic decision-making, thus leading to improved operating performance. Third, to select matching firms with similar ex-ante characteristics as the firm experiencing an activist intervention, we utilize a propensity score matching approach. The set of matching variables used is extensive as it includes pre-activism performance (as in deHaan et al., 2019) and firm size, age, capital expenditures, business risk, financial risk, and research and development expenditures. Finally, we obtain statistically significant results of a reduced negative effect of multiple block share purchasing campaigns targeting the same firm in a particular year. This finding provides important implications for activist coalitions and their targets.

We conduct the empirical study using a sample of 7,563 block share purchase campaigns on 4,178 firms for the period of 1997-2019. Our analysis indicates that the operating performance of targeted firms deteriorates after the block purchase. We find that the results are statistically significant only when the filer is identified as an investment adviser, partnership, corporation, or individual. In addition, the stated purpose for intervention does not affect the corporate operating performance following an activist intervention.

BACKGROUND

After the stock market collapse of 1929, the rights and obligations of shareholders remained uncertain, especially when it came to trading securities of public corporations and management oversight. The Securities and Exchange Commission introduced General SEC Regulation X- 14-7, renumbered 14-8 in 1948.[1] It acknowledged that shareholders had the legal right to communicate with management and other shareholders through the company's proxy material (Marens, 2002). These early shareholder resolutions were put forward by small groups of activist investors and largely ignored by management (Reid and Toffel, 2009).

Individual investors dominated activism from 1942 to the end of the 1970s. The 1980s saw a surge in corporate raiders, where the activist shareholders' objective was to gain control of the firm and replace "poorly performing" management. During this time, some activist investors purchased large blocks of a firm's shares and used their influence to force the firm to make policy decisions designed to increase share prices, such as raising leverage and dividend payouts, reducing R&D and capital expenditure, or increasing share buybacks. Once these initiatives led to share price increases, the investors sold shares. This practice was considered to be problematic as management responded to the demands of myopic short-sighted investors rather than focusing on the long-term success of the company. The establishment of the Council of Institutional Investors in 1985 marked the beginning of activism by institutional investors.

Long-term institutional activism intensified in the 1990s. Pension funds, mutual funds, and labor unions, instead of individual investors, began to acquire large holdings in companies. They aimed to improve operating strategies with the intention of holding on to the shares for a number of years and generating positive returns. Pension funds and mutual funds have to adhere to regulatory and structural constraints, such as limits on the percentage of their portfolio that they can invest in any given firm and restrictions on shorting and investing in illiquid assets. While in the early 1990s pension funds introduced fewer proposals and sought dialogue with management, pension fund activism increased in the latter half of the 1990s when the Department of Labor stated that "active monitoring and communication with corporate management is consistent with a fiduciary's obligations under ERISA" -Department of Labor Interpretative Bulletin 94-2.

In the late 1990s and early 2000s, hedge funds became very active in influencing companies in which they had acquired an ownership stake. Hedge funds did not target operationally distressed firms and did not seek control of the target. The reason for hedge funds to hold an ownership stake in a target company is primarily for hopes of adding to shareholder value. According to PricewaterhouseCoopers "The director's guide to shareholder activism" (2021), hedge funds mostly sponsor activist proposals that are aimed at increasing stock price and shareholder value.

The SEC Securities Act of 1934 defines a "beneficial owner" as any person who directly or indirectly shares voting power or investment power, the power to sell the security (see www.investor.gov for description of Schedules 13D and 13G). When beneficial ownership exceeds five percent of a voting class of equity securities registered under the SEC Act, a person or a group of persons are required to file a Schedule 13D with the SEC. This schedule reports the acquisition and other information within ten days after the purchase. It is filed with the SEC and is provided to the company that issued the securities and each exchange where the security is traded. Depending on circumstances, the filer might be eligible to file abbreviated Schedule 13G. Item 1 of Schedule 13D identifies the filers, item 2 can be checked if a filer is part of a group. Item 4 of the Schedule 13D reports the reason for purchasing securities. The purpose of the acquisition of securities can range from a merger, reorganization or liquidation of the issuer to material changes in business or corporate structure, capitalization, dividend or investment policy, and change in the board of directors. The form reports source and amount of funds used to purchase securities, and the type of the acquirer (broker dealer, bank, insurance company, investment company, investment adviser, employee benefit plan or endowment fund, parent holding company, savings association, church plan, corporation, partnership, individual).

LITERATURE REVIEW

Earlier literature focused on the impact of insider ownership on shareholder returns and firm value. For example, Shleifer and Vishny (1986) show that shareholders with large blocks of shares benefit from the possibility of a hostile takeover and explain the link between corporate governance and firm performance. They find that firm value changes in a concave fashion, increasing with percentage of shares owned, reaching its peak, and then declining when the optimum ownership is reached. Subsequently, Morck et al. (1988) study the effect of managerial ownership on firm valuation as measured by Tobin's Q and find a nonmonotonic relationship. Bethel et al. (1998) study block share purchases of Fortune500 companies for the period of 1980 to 1989 by classifying block share purchases as activist, financial, and strategic. The results indicate that activist block share purchases target highly diversified and poorly performing firms that experience an increase in divestitures, operating profitability and stock price appreciation following the block share purchases.

Demsetz and Villalonga (2001) use a sample of 223 companies from 1976-1980, which is a subsample of the initial Demsetz and Lehn (1985) study. The authors find no statistically significant relation between ownership structure and firm performance. Karpoff (2001), Romano (2001), and Gillan and Starks (2000) investigate the impact of institutional investors and pension funds activity on firm performance, and they report no statistically significant benefits to shareholders after the filing of the Schedule 13D. Chen and Shohfi (2021); Brav et al. (2008, 2015) and Greenwood and Schor (2009) summarize how activist hedge funds target firms. The goals of hedge fund activism, as publicly reported, fall into five categories: maximization of shareholder value, changing capital structure, business strategy, sale of a target company, and enhancing governance.

Coffee and Palia (2015) and Becht et al. (2017) also observe what they call “wolf packs” of funds working together to facilitate the outcomes imposed on target firms. Strine (2016) argues that these wolf packs often have a shorter-term investment strategy which contributes to a corporate culture that is increasingly built on a tradeoff between short-term interests and long-term goals.

Brav et al. (2008) examine 236 hedge funds with 1,032 activist events for the period of 2001 to 2006 and find a 7%-8% increase in the average abnormal returns after the Schedule 13D filing by a hedge fund. Klein and Zur (2009) investigate hostile activist campaigns by hedge funds and find that companies targeted by hedge funds double their dividends, decrease their cash holdings, and increase their debt one year after the initial filing of Schedule 13D. Companies targeted by entrepreneurial investors on the other hand, experience a reduction in R&D and capital expenditures one year after the initial Schedule 13D filing. In both cases, the authors find a significant positive market reaction around the Schedule 13D filing that persists for a year after the filings.

Brav et al. (2022) provide a theoretical framework of the tradeoff between private monitoring costs borne by activists and the public benefits enjoyed by all shareholders. While small blockholding reduces the incentive for intervention as documented in the extant literature, the competition for investor funds leads to a synchronized activism by multiple entities (wolf packs). Anticipated firm value increases in the presence of multiple activists.

Kedia et al. (2021) find that firms with higher levels of activist-friendly institutional ownership are more likely to be targeted by a hedge fund. They demonstrate that hedge fund activism leads to an improved long-term corporate performance in the presence of activist-friendly institutional ownership.

Clifford and Lindsey (2011) examine how activist interventions are related to blockholding and analyst coverage. They study the compensation practices of blockholders to determine the success of monitoring their targets. The authors analyze S&P1500 companies for the period of 1995-2005 and find that the higher the incentive pay for the monitoring organizations, the more effective the monitoring. The authors also demonstrate that active blockholding leads to increased corporate performance as measured by industry-adjusted return on assets.

Edmans et al. (2013) study the effect of liquidity on the type of activism and its effectiveness. Their analysis of 223 hedge funds and their activist actions for the period of 1995 to 2010 demonstrates that the higher the liquidity, the higher the probability of a hedge fund block acquisition. Edmans et al. (2013) find that hedge funds target smaller firms with low market to book value, higher sales growth, higher leverage and more analyst coverage, results that are similar to those obtained by Brav et al. (2008) and Clifford and Lindsey (2011).

By analyzing the text of sell-side reports up to three months before and after the activism intervention, Chen and Shohfi (2021) find that pre-intervention analyst reports are valuable to investors, leading to negative stock market reaction. Their findings suggest that higher activism dictionary in sell-side reports is positively correlated with target stock performance on activism date and predicts activist intervention.

Bebchuk et al. (2015) attempt to answer to the claim that hedge fund activism is short-term oriented and has detrimental effects for the company and shareholders in the long run. The authors utilize the same database as that of Brav et al. (2008) to study hedge fund interventions for the period of 1994-2007. They find that hedge funds target underperforming firms, and the operating performance of the firms, as measured by ROA and Tobin’s Q, improves up to five years after the hedge fund intervention. This holds true even when hedge fund activists are openly adversarial and also engage in practices that might reduce a firm’s future investment resources through increased leverage, increased dividend payout, or decreased R&D expenditures. However, Cremers et al. (2020) argue that the Bebchuk et al. (2015) study suffers from selection bias since activists target firms with lagging performance compared to their peers. The authors replicate the Bebchuk et al. (2015) study, using a nearest neighbor matching approach to create a control sample, and find that the value of firms targeted by activists improves by less than those not subject to activism. This result indicates that hedge fund activism alone cannot account for an increase in firm value following an activist campaign.

DeHaan et al. (2019) measure both short-term abnormal stock returns in a 21-day window surrounding the activist intervention and long-term operating performance for the targets. They find that equal-weighted mean returns are significantly positive at 5.4%, and the cumulative pre- and post-activism equal-weighted mean one-year and two-year returns are significantly positive at 6.8% and 5.9% respectively. However, the smallest 20% of the target companies contribute the most to positive returns, and the largest 80% show positive but insignificant returns within three months of activism.

Brick et al. (2019) attribute improved operating performance of targeted firms to hedge fund human capital. In particular, the authors use hedge fund acquisitions from 1994 to 2011 (same data set as that of Brav et al., 2015) and demonstrate that firms targeted by hedge funds with experience in a particular industry have higher abnormal returns than peer firms for up to three years after the initial activist acquisition.

Bebchuk et al. (2020) show that settlement agreements between activist hedge funds and target companies are more likely when the activist has a credible threat to win board seats in a proxy fight. They document that settlements focus on boardroom changes resulting in increases in CEO turnover, rises in shareholder payouts, and the greater likelihood of a sale or a going-private transaction. They find no evidence that such actions enable activists to extract significant rents at the expense of other investors.

Wang and Wu (2020) demonstrate that hedge fund activism is triggered by negative media coverage of a target. Firms with news related to bad acquisitions, lower earnings, and/or lower credit ratings, have a higher probability of being targeted by a hedge fund. The authors also show increased short-term and long-term (three year) shareholder returns of the targeted firms.

HYPOTHESIS, DATA AND METHODOLOGY

This study extends the literature on block share purchases in that we investigate its effect on long-term firm performance for up to five years after the intervention (Brav et al., 2008; Clifford and Lindsey, 2011; deHaan et al., 2019). We focus on blockholder interventions by different types of filers as reported in the SEC Schedule 13D during the 1998-2019 period, and firm performance of the targets for three years before and five years after the interventions. It extends the work by Bebchuk et al. (2015) as we explore the effect of blockholding by different types of investors and different purposes as listed in Item 4 of Schedule 13D. Moreover, to select matching firms with similar ex-ante characteristics as the firm experiencing an intervention, we utilize a propensity score matching approach. The set of matching variables includes pre-activism performance (as in deHaan et al., 2019) and firm size, age, capital expenditures, business risk, financial risk, and research and development expenditures. We also examine multiple block purchases on the same target and find them beneficial as they lead to a better operating performance than that of firms experiencing only one activist intervention.

DATA SOURCES

Block share purchase filings can be found in the Securities and Exchange Commission (SEC) Schedule 13D. The SEC Act (1934) requires that investors acquiring more than 5% of a company's stock file the 13D form with the SEC to report the share purchase and other information, including purpose of the purchase, within 10 days after the transaction. Filers must also disclose their identities in Item 2, and the reason for acquiring the shares in Item 4 of Schedule 13D – "Purpose of Transaction." In our sample, the purpose of block share purchases is classified as investment, governance, and business strategy. We use keywords to distinguish between the categories. Some filers may report multiple purposes. The filer is coded by type (item 2 of Schedule 13D) as individual, partnership, investment advisor, institutional investor, private equity company or an asset management company. The date of filing is also recorded. Targets are identified and their SIC 4-digit numbers are noted. The sample is screened for the availability of appropriate data, including SIC codes and annual reports over a 5-year period before and after the filing date. We eliminate all firms for which no SIC code was listed and for which subsequent searches did not provide any information on their SIC codes.

BLOCK SHARE PURCHASES

Table 1 provides summary statistics at the intervention level. Our sample spans from 1997 to 2019, resulting in 7,563 activist campaigns.

Panel A of Table 1 presents the percentage of the target that is acquired, and the purpose of the acquisition as stated in item 4 of Schedule 13D, which we code as governance, business strategy or investment. The average percentage of shares acquired is 16.2%, with a median of 9.6%. We also find that 8.01% of filers in our sample reported governance as the purpose of intervention, 40.3% reported business strategy as its purpose, and 69.8% reported that intervention was due to investment reasons. These categories overlap since some activists declared more than one purpose of their intervention.

Panel B of Table 1 breaks down our block purchase campaign sample by category of filer. The largest percentage of Schedules 13D was filed by individuals (28.6%), followed by partnerships (14.3%), corporations (13.5%), and other filers.

The source of activist campaign funding is reported in Panel C of Table 1. The largest reported sources are working capital of the filer (17.1%), funds of the affiliate of the filer (9.03%), bank capital (0.608%), and funds of the subject company that is being acquired (0.502%).

Panel D of Table 1 presents the summary statistics for repeated interventions. It is worth exploring the sample for multiple interventions either by the same filer, or for the same target. Panel D shows that the total maximum number of campaigns for the same target was 22 interventions, with the mean of 3.45 total campaigns (variable is denoted by Total Campaigns on Target). Total Campaigns by Activist is the total number of interventions per the same acquirer, with the mean of 3.4 campaigns and the maximum of 39 campaigns. Annual Campaigns on Target is the number of interventions per target per year, where on average 1.66 campaigns were reported, with the maximum number of 22 campaigns. Finally, Annual Campaigns on Target by Activist is the number of interventions per target done by the same acquirer. It is 1.33 campaigns on average, with a maximum of 10 campaigns.

We do not observe formal coordination of actions by multiple activists because that would require they file Form 13D as a group, declare their intentions (Item 4), and risk potential lawsuits by target management. Moreover, filing as a group may trigger poison pills that could restrict group holdings.

Table 1: Summary Statistics at the Intervention Level.

Panel A: Percentage acquired and purpose	**Mean, %**
Percentage acquired	16.2
Purpose governance	8.01
Purpose business strategy	40.3
Purpose investment	69.8
Panel B: Category of filer	**Mean, %**
Bank	0.17
Broker dealer	0.55
Company	13.5
Employee benefit plan	0.38
Parent holding company	3.23
Investment advisor	5.73
Insurance company	0.13
Individual	28.6
Investment company	0.36
Limited liability company	1.03
Partnership	14.3
Savings association	0.01
Other	32.01

Panel C: Source of financing	Mean, %
Affiliate	9.0
Bank	60.8
Subject company	0.50
Working capital	17.1
Other	12.57

Panel D: Multiple interventions	mean	min	median	max
Total Campaigns on Target	3.45	1	3	22
Total Campaigns by Activist	3.4	1	1	39
Annual Campaigns on Target	1.66	1	1	22
Annual Campaigns on Target by Activist	1.33	1	1	10

Number of observations is 7,563.

TARGETS OF BLOCK PURCHASE INTERVENTIONS

Table 2 describes our sample of target firms. We collected full set of variables for 4,178 firms targeted by blockholding investors during the 1997-2019 period. The financial characteristics of firms are obtained from Compustat. Corporate performance and firm profitability are measured by ROA and by Tobin's Q.

Annual values of ROA and Tobin's Q are calculated for individual target firms for up to three years before the activist intervention and up to five years after the intervention. Each firm's industry is identified by a four-digit SIC code, and the average ROA and Tobin's Q are calculated for each industry. This step facilitates the calculation of industry adjusted ROA and Tobin's Q that are used in the regression analysis. Industry adjusted performance metrics are appropriate since we want to examine the firm performance in relation to their peers. A positive ROA or Tobin's Q indicates that the targeted firm outperforms its peers. The size of firms measured by natural logarithm of Total Assets and firm Age are used as control variables.

Panel A of Table 2 reveals that activist intervention targets have the average Tobin's Q of 1.36, mean Return on Assets of 3.43% and average Total Assets of $1,964 million. Panel A reports firm fundamentals as of the year of intervention. In comparison, Panel B of Table 2 reports the average statistics for all 113,637 Compustat firms that did not have intervention during the same time period. This suggests that firms without activist campaign had a higher mean Tobin's Q of 1.52, a higher mean ROA of 6.03%, and higher Total

Assets of $6,829 million. Therefore, on average, activist targets are smaller in size and less profitable before the intervention.

Panel C of Table 2 reports the summary statistic of a propensity-score matched sample of industry peers. The purpose of propensity score matching is to select a firm with similar ex-ante firm characteristics as the firm experiencing an activist intervention. To implement this method, the propensity score for each firm is first estimated. In essence, the model estimates the probability that a given firm with certain characteristics is experiencing an intervention (Dehejia and Wahba, 2002). The propensity score is the probability of being a firm in the new member state conditional on x, *p(x) = probability (D = 1—x)*, where D=1 for firms that experienced an intervention and D=0 for other firms. The probability is composed using firm characteristics within a logistic regression framework. We matched the firms based on lagged Total Assets, Tobin's Q, Capital Expenditures, R&D Expense, among other firm characteristics.

The results of propensity score matching show that the average Tobin's Q of the control firms without the intervention is 1.44, which is higher than the average Tobin's Q for intervention targets. The difference is statistically significant at the 1% confidence level. The same result holds for ROA and it is significant at the 10% confidence level. Book Leverage is also statistically different for intervention targets and control firms, with a 5% confidence level. This finding implies lower profitability and higher leverage of target firms since block share investors tend to approach underperforming companies.

The matching was performed for the year of the intervention and the table reports firm fundamentals for that year. Also, the matching was done with replacement, meaning that the control could be selected more than once. Matching with replacement minimizes the propensity score distance between the matched units since each treated unit is matched to the closest control, even if the control has been selected before. This resulted in a smaller sample of control firms when compared to treated firms in Panel A.

Table 2: Summary Statistics at the Firm-Year Level.

Panel A: Target firms (Intervention = 1)	**mean**	**st.dev**	**min**	**median**	**max**
Tobin's Q	1.36	1.24	0.112	1	11.1
Return on Assets	3.43	27.4	-4.31	7.29	64.6
Assets - Total	1,964	27,048	1.04	193	1,660,763
Log of Total Assets	5.34	1.99	3.73	5.26	14.3
R&D Expense	4.35	9.28	0	0	92.9

Capital Expenditures	74.5	441	-295	4.75	14,742
Company Age	19	14.3	2	15	70
Book Leverage	24.2	22.5	0	19.1	99.8
Operating Leverage	1.06	82.0	2.09	90.3	4.52
Observations	4,178				

Panel B: Control firms (Intervention = 0)	**mean**	**st.dev**	**min**	**median**	**max**	**t-stat**
Tobin's Q	1.52	1.42	.111	1.08	11.7	6.869***
Return on Assets	6.03	.267	-5.12	9.58	0.649	6.192***
Assets - Total	6,829	59,918	1	359	3510975	5.229***
Log of Total Assets	5.89	2.34	0.10	5.88	15.1	15.198***
R&D Expense	4.25	9.28	0	0	93.7	-0.725
Capital Expenditures	221	1,332	-994	9.18	50,234	7.098***
Company Age	17.2	14.1	2	13	71	-7.749***
Book Leverage	22.4	21.0	0	18.1	1	-5.341***
Operating Leverage	91.2	77.0	2.01	73.9	4.52	-12.216***
Observations	113,637					

Panel C: Propensity score matched controls	**mean**	**st.dev**	**min**	**median**	**max**	**t-stat**
Tobin's Q	1.44	1.32	0.113	1.04	11	4.549***
Return on Assets	5.00	24.5	-4.93	8.59	64.6	2.056*
Assets - Total	3,532	53,89	1.07	220	3,065,5	1.458

		1			53	
Log of Total Assets	5.41	2.23	0.068	5.39	14.9	0.235
R&D Expense	4.46	9.52	0	0	89.8	0.713
Capital Expenditures	85	440	0	5.83	14,678	0.693
Company Age	19.9	15.1	3	15	70	-0.100
Book Leverage	22.5	21.4	0	17.7	99.9	-2.688**
Operating Leverage	106	.83	2.03	88.2	451	0.304
Observations	3,640					

Tobin's Q is calculated using Chung and Pruitt (1994) approach: ($PRCC_F \times CSHO + PF + DLTT + DLC)/AT$. To calculate the book value of preferred stock (PF), we use the redemption (PSTKRV), liquidation (PSTKL), or par value (PSTK), in that order, depending on availability. Return on assets (ROA) is operating income before depreciation scaled by the lagged book value of assets (OIBDP/L.AT). R&D is the research and development expense scaled by the lagged book value of assets (XRD/L.AT). Book leverage is calculated as the sum of short-term debt and long-term debt divided by the book value of assets (DLC+DLTT)/AT. Operating leverage is operating costs is cost of goods sold plus selling, general, and administrative expenses scaled by the total assets (COGS+XSGA)/AT. The t-statistics of the differences between control and target firms' average values are reported in the last column of Panels B and C.

LONG-TERM DYNAMICS OF CORPORATE PROFITABILITY

Table 3 reports the results of panel regressions of Tobin's Q and ROA using the propensity score-matched sample. The sample size in Columns 1 and 2 is different from Columns 3 and 4, which is due to the fact that we are using a different dependent variable which affects the number of observations. We use industry-adjusted Tobin's Q (IATQ) and industry-adjusted ROA (IAROA) to control for the possibility that regression measurement errors are industry specific. Industry-adjusted performance measures are calculated by comparing firm's Tobin's Q (ROA) with the yearly median of all the companies available in the same industry as the sample firm. Tobin's Q is calculated using Chung and Pruitt (1994) approach: ($PRCC_F \times CSHO + PF + DLTT + DLC)/AT$. To calculate the book value of preferred stock (PF), we use the redemption (PSTKRV), liquidation (PSTKL), or par value (PSTK), in that order, depending on availability. Return on assets (ROA) is operating income before depreciation scaled by the lagged book value of assets (OIBDP/L.AT). R&D is the research and development expense scaled by the lagged book value of assets (XRD/L.AT). Book leverage is calculated as the sum of short-term debt and long-term debt divided by the book value of assets (DLC+DLTT)/AT. Operating leverage is operating costs is cost of goods sold plus selling, general, and administrative expenses scaled by the total assets (COGS+XSGA)/AT.

Our findings indicate that firms targeted by activist investors have lower performance as indicated by the negative coefficient for Tobin's Q and industry-adjusted Tobin's Q for up to two years after the intervention. This runs contrary to the findings by Bebchuk et al. (2015) and deHaan et al. (2019). The results are statistically significant at the 1% level for one year after the intervention and at 5% confidence level for two years after the activist campaign. When ROA and industry-adjusted ROA are used as measures of operating performance, the results are statistically significant only for one year after the intervention.

Table 3: Dynamics of Tobin's Q and ROA.

	Tobin's Q (1)	**Industry-adjusted Tobin's Q (2)**	**Return on assets (3)**	**Industry-adjusted ROA (4)**
(t-3)	-0.0087	-0.0165	-0.0064	-0.0059
	(-0.43)	(-0.84)	(-1.55)	(-1.45)
(t-2)	-0.0618***	-0.0572***	-0.0089***	-0.0104***
	(-3.74)	(-3.65)	(-2.72)	(-3.18)
(t-1)	-0.0737***	-0.0618***	-0.0128***	-0.0142***
	(-4.88)	(-4.38)	(-3.66)	(-4.34)
Block Purchase	-0.0916***	-0.0832***	-0.0185***	-0.0169***
	(-5.45)	(-5.40)	(-5.35)	(-5.23)
(t+1)	-0.1410***	-0.1410***	-0.0169***	-0.0153***
	(-9.29)	(-9.79)	(-4.92)	(-4.69)
(t+2)	-0.0624***	-0.0617***	-0.0111***	-0.0135***
	(-3.74)	(-3.85)	(-3.12)	(-3.93)
(t+3)	-0.0692***	-0.0687***	-0.0026	-0.0012
	(-4.11)	(-4.32)	(-0.78)	(-0.38)
(t+4)	-0.0273	-0.0426**	0.0005	-0.0013

	(-1.44)	(-2.36)	(0.15)	(-0.37)
(t+5)	-0.0437*	-0.0505**	0.0032	0.0024
	(-1.89)	(-2.29)	(0.73)	(0.57)
Log of Total Assets	0.0164*	0.0198**	0.0281***	0.0251***
	(1.96)	(2.54)	(21.93)	(19.46)
R&D Expense	3.9930***	3.2360***	-0.9150***	-0.759***
	(19.09)	(16.02)	(-16.70)	(-14.50)
Capital Expenditures	<-0.0001	<-0.0001	<-0.0001***	<-0.0001***
	(-0.04)	(-0.46)	(-4.81)	(-5.44)
Company Age	-0.0005***	-0.0037***	0.0002*	<-0.0001
	(-5.47)	(-4.52)	(1.69)	(-0.28)
Book Leverage	-0.4660***	-0.2930***	-0.1030***	-0.110***
	(-7.56)	(-5.24)	(-11.93)	(-12.52)
Operating Leverage	-0.0545**	-0.0241	-0.0081***	-0.0122***
	(-2.53)	(-1.26)	(-2.67)	(-4.01)
Constant	1.9070***	0.1680	0.0544***	-0.0183
	(8.45)	(0.92)	(2.67)	(-1.09)
Year Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes

Observations	56,234	56,234	57,464	57,464
Adjusted R^2	0.211	0.052	0.222	0.151

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If a coefficient is very small, i.e., less than 0.0001, then it is reported as <0.0001.

Panel Regression is performed using the propensity score-matched sample. This table reports dynamics of Tobin's Q and ROA from three years before (-3) to five years after (5+) a Block Share Purchase for target and control firms. The variable *t* implies the period around intervention.

Table 4 reports panel regression of firm performance for target and control firms using the propensity-score matched sample. Indicator variable POSTxPURCHASE refers to the companies that experienced a block purchase and exist five years after the intervention. It is similar to the approach used by An and Wu (2021). To define the post-intervention period, we use five years after the interventions. POST*PURCHASE is an interaction dummy variable equal to one if the firm-year observation is within t to t+5 years of a block purchase event (for target firms) or a pseudo-event year (for matched firms). The results confirm the deteriorating operating performance of targets after the intervention, based on statistically significant negative coefficient of POST*PURCHASE variable in regressions on Tobin's Q, industry-adjusted Tobin's Q, ROA and industry-adjusted ROA. Defining the post-intervention period as t+3 or t+4 yields similar results (the results are not reported for brevity).

Table 4: Panel Regression

	TQ	**IATQ**	**ROA**	**IAROA**
	(1)	**(2)**	**(3)**	**(4)**
Post x Purchase	-0.1350***	-0.1450***	-0.0143***	-0.0139***
	(-6.19)	(-7.06)	(-3.92)	(-3.87)
Log of Total Assets	0.0191**	0.0223***	0.0286***	0.0256***
	(2.30)	(2.87)	(22.23)	(19.81)
R&D Expense	3.9900***	3.2340***	-0.9150***	-0.759***
	(19.05)	(15.98)	(-16.69)	(-14.49)
Capital Expenditures	<-0.0001	<-0.0001	<-0.0001***	<-0.0001***
	(-0.01)	(-0.42)	(-4.85)	(-5.53)

Company Age	-0.0049***	-0.0037***	0.0002*	<-0.0001
	(-5.48)	(-4.53)	(1.68)	(-0.30)
Book Leverage	-0.4860***	-0.3110***	-0.1060***	-0.113***
	(-7.87)	(-5.57)	(-12.24)	(-12.82)
Operating Leverage	-0.0560***	-0.0255	-0.0083***	-0.0124***
	(-2.59)	(-1.33)	(-2.74)	(-4.08)
Constant	1.883***	0.149	0.0506**	-0.0222
	(8.34)	(0.82)	(2.49)	(-1.35)
Year Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Observations	56,234	56,234	57,464	57,464
Adjusted R^2	0.209	0.051	0.220	0.150

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If a coefficient is very small, i.e., less than 0.0001, then it is reported as <0.0001.

Table 5 reports panel regression of firm performance for target and control firms with several interaction variables. POSTxPURCHASE refers to the companies that experienced a block purchase and exist five years after the intervention. The purpose of the intervention as stated in item 4 of Regulation 13D is coded as governance, strategy or investment. We use interaction terms between POST*PURCHASE variable and each of the three purpose variables. The results indicate that target firms approached by activists for the purpose of changing the company's strategy experienced an even more negative drop in Tobin's Q after intervention. The same result holds for firms that were targeted for the purpose of investment. Interaction of POST*PURCHASE and governance variables yields no significant results for target profitability after the block- purchase campaign.

Table 5: Panel Regression. Purpose

	TQ (1)	ROA (2)	TQ (3)	ROA (4)	TQ (5)	ROA (6)
Post * Purchase	-0.1420***	-0.0141***	-0.1240***	-0.0120***	-0.1190***	-0.0108***
	(-6.62)	(-3.86)	(-5.58)	(-3.39)	(-5.38)	(-3.10)
Post * Governance	-0.1280*	-0.0162				
	(-1.88)	(-1.09)				
Post * Strategy			-0.1440***	-0.0316*		
			(-3.55)	(-1.89)		
Post * Investment					-0.1260***	-0.0282**
					(-3.51)	(-2.39)
Log of Total Assets	0.0164**	0.0286***	0.0189**	0.0285***	0.0188**	0.0285***
	(2.00)	(22.23)	(2.27)	(22.29)	(2.25)	(22.26)
R&D Expense	4.0210***	-0.9150***	3.990***	-0.9150***	3.9910***	-0.9150***
	(19.15)	(-16.69)	(19.05)	(-16.69)	(19.06)	(-16.70)
Capital Expenditures	<0.0001	<-0.0001***	<-0.0001	<-0.0001***	<-0.0001	<-0.0001***
	(0.06)	(-4.85)	(-0.02)	(-4.85)	(-0.02)	(-4.85)
Company Age	-0.0050***	0.0002*	-0.0049***	0.0002*	-0.0049***	0.0002*
	(-5.85)	(1.68)	(-5.49)	(1.68)	(-5.49)	(1.67)
Book Leverage	-0.4720***	-0.1060***	-0.4840***	-0.1060***	-0.4840***	-0.1060***
	(-7.75)	(-12.24)	(-7.85)	(-12.27)	(-7.84)	(-12.22)
Operating Leverage	-0.0620***	-0.0083***	-0.0559***	-0.0083***	-0.0562***	-0.0083***
	(-2.88)	(-2.74)	(-2.59)	(-2.74)	(-2.60)	(-2.76)
Constant	1.6050***	0.0505**	1.8850***	0.0509**	1.8830***	0.0505**
	(7.09)	(2.48)	(8.35)	(2.50)	(8.34)	(2.49)

Year Fixed Effects	No	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,234	57,464	56,234	57,464	56,234	57,464
Adjusted R^2	0.191	0.220	0.209	0.221	0.209	0.221

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If a coefficient is very small, i.e., less than 0.0001, then it is reported as <0.0001.

Table 6 takes the category of the filer into consideration. This table reports panel regression of firm performance for target and control firms using the propensity-score matched sample. POSTxPURCHASE refers to the companies that experienced a block purchase and exist five years after the intervention. Interaction terms between Post and four main categories of acquirers (partnership (PN), company (CO), individual (IN), and investment advisor (IA) are included. There is a negative and statistically significant difference in the performance of the companies that experienced interventions and the control sample when the filer is identified as a partnership, an investment adviser, an individual, or a company. This result indicates that the profitability of firms targeted by these four categories of blockholders decreased to a higher degree.

Table 6: Panel Regression. Category of acquirer.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	TQ	ROA	TQ	ROA	TQ	ROA	TQ	ROA
Post * Purchase	-0.1300***	-0.0131***	-0.1310***	-0.0129***	-0.1280***	-0.0128***	-0.1320***	-0.0142***
	(-5.94)	(-3.60)	(-5.99)	(-3.60)	(-5.81)	(-3.63)	(-6.06)	(-3.88)
Post x PN	-0.1770***	-0.0427***						
	(-2.71)	(-3.29)						
Post x CO			-0.1580***	-0.0576***				
			(-2.77)	(-3.26)				
Post x IN					-0.1330***	-0.0297*		
					(-3.07)	(-1.70)		
Post x IA							-0.208***	-0.0094
							(-2.88)	(-0.78)

Log of Total Assets	0.0191**	0.0286***	0.0191**	0.0286***	0.0190**	0.0286***	0.0192**	0.0286***
	(2.30)	(22.23)	(2.30)	(22.24)	(2.28)	(22.24)	(2.30)	(22.23)
R&D Expense	3.992***	-0.9150***	3.990***	-0.9150***	3.992***	-0.9150***	3.990***	-0.9150***
	(19.06)	(-16.69)	(19.06)	(-16.69)	(19.07)	(-16.68)	(19.06)	(-16.69)
Capital Expenditures	<-0.0001	<-0.0001***	<-0.0001	<-0.0001***	<-0.0001	<-0.0001***	<-0.0001	<-0.0001***
	(-0.02)	(-4.85)	(-0.02)	(-4.86)	(-0.02)	(-4.85)	(-0.02)	(-4.85)
Company Age	-0.0049***	0.0002*	-0.0049***	0.0002*	-0.0049***	0.0002*	-0.0049***	0.0002*
	(-5.49)	(1.67)	(-5.49)	(1.66)	(-5.48)	(1.68)	(-5.47)	(1.68)
Book Leverage	-0.4850***	-0.1060***	-0.4860***	-0.1060***	-0.4850***	-0.1060***	-0.4850***	-0.1060***
	(-7.87)	(-12.24)	(-7.87)	(-12.24)	(-7.86)	(-12.25)	(-7.87)	(-12.24)
Operating Leverage	-0.0559***	-0.0083***	-0.0560***	-0.0083***	-0.056***	-0.0083***	-0.056***	-0.0083***
	(-2.59)	(-2.74)	(-2.60)	(-2.75)	(-2.60)	(-2.74)	(-2.59)	(-2.74)
Constant	1.883***	0.0506**	1.8840***	0.0508**	1.8840***	0.0507**	1.8820***	0.0505**
	(8.36)	(2.48)	(8.36)	(2.49)	(8.36)	(2.49)	(8.34)	(2.48)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56234	57464	56234	57464	56234	57464	56234	57464
Adjusted R^2	0.209	0.221	0.209	0.221	0.209	0.221	0.209	0.220

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If a coefficient is very small, i.e., less than 0.0001, then it is reported as <0.0001.

Another important result is reported Table 7. This table reports panel regression of firm performance for target and control firms using the propensity-score matched sample. POSTxPURCHASE refers to the companies that experienced a block purchase and exist five years after the intervention. Multiple indicates multiple activist interventions for a given target in a given year. We examine multiple block share purchases for a certain company in a given year. Specifications (1) and (2) of Table 7 show a positive and significant coefficient for *POST* Multiple* interaction dummy in the Tobin's Q regressions. This implies that there is less of a drop in profitability of targets after they experience multiple interventions. Similar to prior research documenting a beneficial effect of activist ”wolfpacks” on shareholders (Becht et al. (2017) and Strine (2016)) we find that multiple block share purchasing campaigns improve profitability of a target firm. Becht et al. (2017) estimate that ”wolf packs are associated with almost one quarter of all engagements” and show that ”they achieve some of the highest returns for target shareholders” (p. 2934).

Table 7: Panel Regression propensity score matched sample

	TQ (1)	**IATQ (2)**	**ROA (3)**	**IAROA (4)**
Post x Purchase	-0.1680***	-0.1580***	-0.0165***	-0.0173***
	(-6.52)	(-6. 61)	(-3.49)	(-3.82)
Post x Multiple	0.0461*	0.0191	0.0003	0.0047
	(1.88)	(0.84)	(0.69)	(1.13)
Log of Total Assets	0.0193**	0.0224***	0.0286***	0.0256***
	(2.31)	(2.88)	(22.23)	(19.82)
R&D Expense	3.9870***	3.2330***	-0.9150***	-0.759***
	(19.04)	(15.97)	(-16.69)	(-14.50)
Capital Expenditures	<-0.0001	<-0.0001	<-0.0001***	<-0.0001***
	(-0.02)	(-0.43)	(-4.85)	(-5.53)
Company Age	-0.0049***	-0.0037***	0.0002*	< -0.0001
	(-5.49)	(-4.53)	(1.67)	(-0.31)
Book Leverage	-0.4860***	-0.3110***	-0.1060***	-0.113***
	(-7.87)	(-5.57)	(-12.24)	(-12.82)
Operating Leverage	-0.0558***	-0.0255	-0.0083***	-0.0124***

	(-2.59)	(-1.33)	(-2.74)	(-4.08)
Constant	1.8750***	0.145	0.0500**	-0.0231
	(8.33)	(0.80)	(2.45)	(-1.40)
Year Fixed Effects	Yes	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes	Yes
Observations	56234	56234	57464	57464
Adjusted R^2	0.209	0.051	0.220	0.150

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. If a coefficient is very small, i.e., less than 0.0001, then it is reported as <0.0001.

CONCLUSION

This paper studies the long-term effects of shareholder block share purchases on a firm's performance. We use the sample of investors who filed 13D forms with the SEC. The main reason for active participation by investors in the monitoring of corporations is the potential to advance their agenda and improve firm value.

Our results indicate that there is a deterioration in corporate performance, particularly when the filer is identified as an investment advisor, a corporation, a partnership, or an individual. Our findings may be partially explained by the fact that our sample consists of larger-size target firms. The literature finds that the positive effect of activism is limited to smaller targets, with the smallest 20% of the target companies contributing the most to positive returns, as in deHaan et al. (2019). It may also be the case that management of target firms is forced to devote corporate resources to fighting activists thus impacting negatively the operational performance of their firm. Additionally, the presence of blockholders may be perceived as a threat of increased costs if the investors decide to withdraw.

Another interesting finding is the positive influence of multiple blockholders on Tobin's Q of a target. We posit that when several blockholders target the same company in a given year, it sends a positive signal to the market participants and results in a lower profitability loss.

ENDNOTES

[1] Rule 14a-8 (b, c, i) of the Securities and Exchange Act of 1934 that regulates shareholder proposal rules was amended in September 2020. The amount and the length of time of ownership of the proposing shareholder for initial proposal and the requirements for resubmission were modified. https://www.sec.gov/news/press-release/2020-220; https://corpgov.law.harvard.edu/2020/10/12/sec-

increases-rule-14a-8-thresholds/

REFERENCES

An, H., L. L. and Q. Wu (2021). The impact of hedge fund activism on corporate cost behavior. *Working paper, presented at the 2021 FMA in Denver, Colorado*.

Barber, B. M. (2007). Monitoring the monitor: Evaluating CalPERS ’activism. *The Journal of Investing 16* (4), 66–80.

Bebchuk, L. A., A. Brav, and W. Jiang (2015). The long-term effects of hedge fund activism. *Columbia Law Review* 115 (5), 1085–1155.

Bebchuk, L. A., A. Brav, W. Jiang, and T. Keusch (2020). Dancing with activists. *Journal of Financial Economics 137* (1), 1–41.

Becht, M., J. Franks, J. Grant, and H. F. Wagner (2017). Returns to hedge fund activism: An international study. *The Review of Financial Studies 30* (9), 2933–2971.

Bethel, J. E., J. P. Liebeskind, and T. Opler (1998). Block share purchases and corporate performance. *The Journal of Finance 53* (2), 605–634.

Brav, A., A. Dasgupta, and R. Mathews (2022). Wolf pack activism. *Management Science 68* (8), 5557–5568.

Brav, A., W. Jiang, and H. Kim (2015). Recent advances in research on hedge fund activism: Value creation and identification. *Annual Review of Financial Economics 7*, 579–595.

Brav, A., W. Jiang, F. Partnoy, and R. Thomas (2008). Hedge fund activism, corporate governance, and firm performance. *The Journal of Finance 63* (4), 1729–1775.

Brick, I. E., Y. Chen, J.-K. Kang, and J.-M. Kim (2019). Does hedge fund managers’ industry experience matter for hedge fund activism? *Available at SSRN 3407718*.

Brown, L. D., A. C. Call, M. B. Clement, and N. Y. Sharp (2016). The activities of buy-side analysts and the determinants of their stock recommendations. *Journal of Accounting and Economics 62* (1), 139–156.

Caton, G. L., J. Goh, and J. Donaldson (2001). The effectiveness of institutional activism. *Financial Analysts Journal 57* (4), 21–26.

Chen, H. A. and T. Shohfi (2021). Do sell-side analysts play a role in hedge fund activism? Evidence from textual analysis. *Contemporary Accounting Research 39 (3),* 1583-1614.

Clifford, C. and L. Lindsey (2011). Blockholder compensation, activism. Technical report, Working Paper, University of Kentucky.

Coffee, J. C. J. and D. Palia (2015). The wolf at the door: The impact of hedge fund activism on corporate governance. *Journal of Corporation Law 41*, 545.

Cremers, M., E. Giambona, S. M. Sepe, and Y. Wang (2020). Hedge fund activism and long-term firm value. *Working Paper, Yale University*.

deHaan, E., D. Larcker, and C. McClure (2019). Long-term economic consequences of hedge fund activist interventions. *Review of Accounting Studies 24* (2), 536–569.

Dehejia, R. H. and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics 84* (1), 151–161.

Del Guercio, D. (1996). The distorting effect of the prudent-man laws on institutional equity investments. *Journal of Financial Economics 40* (1), 31–62.

Del Guercio, D. and J. Hawkins (1999). The motivation and impact of pension fund activism.

Journal of Financial Economics 52 (3), 293–340.

Demsetz, H. and K. Lehn (1985). The structure of corporate ownership: Causes and consequences. *Journal of Political Economy 93* (6), 1155–1177.

Demsetz, H. and B. Villalonga (2001). Ownership structure and corporate performance. *Journal of Corporate Finance 7* (3), 209–233.

Denes, M. R., J. M. Karpoff, and V. B. McWilliams (2017). Thirty years of shareholder activism: A survey of empirical research. *Journal of Corporate Finance 44*, 405–424.

Edmans, A., V. W. Fang, and E. Zur (2013). The effect of liquidity on governance. *The Review of Financial Studies 26* (6), 1443–1482.

English, P. C. I., T. I. Smythe, and C. R. McNeil (2004). The "CalPERS effect" revisited. *Journal of Corporate Finance 10* (1), 157–174.

Francis, B. B., I. Hasan, Y. V. Shen, and Q. Wu (2021). Do activist hedge funds target female CEOs? The role of CEO gender in hedge fund activism. *Journal of Financial Economics 141* (1), 372–393.

Gillan, S. L. (1995). *Shareholder activism through the proxy mechanism: An empirical investigation*. Ph. D. thesis, The University of Texas at Austin.

Gillan, S. L. and L. T. Starks (2000). Corporate governance proposals and shareholder activism: The role of institutional investors. *Journal of financial Economics 57* (2), 275– 305.

Gillan, S. L. and L. T. Starks (2009). The evolution of shareholder activism in the United States. In *US Corporate Governance*, pp. 202–240. Columbia University Press.

Greenwood, R. and M. Schor (2009). Investor activism and takeovers. *Journal of Financial Economics 92* (3), 362–375.

Karpoff, J. M. (2001). The impact of shareholder activism on target companies: A survey of empirical findings. *Available at SSRN 885365.*

Karpoff, J. M., P. H. Malatesta, and R. A. Walkling (1996). Corporate governance and shareholder initiatives: Empirical evidence. *Journal of Financial Economics 42* (3), 365– 395.

Kedia, S., L. T. Starks, and X. Wang (2021). Institutional Investors and Hedge Fund Activism. *The Review of Corporate Finance Studies 10* (1), 1–43.

Klein, A. and E. Zur (2009). Entrepreneurial shareholder activism: Hedge funds and other private investors. *The Journal of Finance 64* (1), 187–229.

Marens, R. (2002). Inventing corporate governance: The mid-century emergence of shareholder activism. *Journal of Business & Management 8* (4), 365- 389.

Morck, R., A. Shleifer, and R. W. Vishny (1988). Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics 20*, 293–315.

Nesbitt, S. L. (1994). Long-term rewards from shareholder activism: a study of the ”CalPERS” effect. *Journal of Applied Corporate Finance 6* (4), 75–80.

Prevost, A. K. and R. P. Rao (2000). Of what value are shareholder proposals sponsored by public pension funds. *The Journal of Business 73* (2), 177–204.

Reid, E. M. and M. W. Toffel (2009). Responding to public and private politics: Corporate disclosure of climate change strategies. *Strategic Management Journal 30* (11), 1157–1178.

Romano, R. (2001). Less is more: making institutional investor activism a valuable mechanism of corporate governance. *Yale J. on Reg. 18*, 174.

Shleifer, A. and R. W. Vishny (1986). Large shareholders and corporate control. *Journal of Political Economy 94*, 461 – 488.

Smith, M. P. (1996). Shareholder activism by institutional investors: Evidence from CalPERS. *The Journal of Finance 51* (1), 227–252.

Song, W.-L. and S. H. Szewczyk (2003). Does coordinated institutional investor activism reverse the fortunes of underperforming firms? *Journal of Financial and Quantitative Analysis 38* (2), 317–336.

Strine, L. E. J. (2016). Who bleeds when the wolves bite: A flesh-and-blood perspective on hedge fund activism and our strange corporate governance system? *Yale Law Journal 126 (6)*, 1870.

Wahal, S. (1996). Pension fund activism and firm performance. *Journal of Financial and Quantitative Analysis 31* (1), 1–23.

Wang, X. and G. Wu (2020). The role of the media in hedge fund activism. *Available at SSRN 3584071.*

Zenner, M., A. Shivdasani, D. Darius, A. Castaldo, and T. Sullivan (2005). Hedge funds at the gate. *Citigroup Global Markets*.

REFEREE ACKNOWLEDGMENT

The New York State Economics Association is grateful for the time spent by the following referees in recent years. Without your efforts, we would not be able to produce this journal.

1. Faisal Abbas
2. Renzo Alvarez
3. James T. Bang
4. Supravat Bagli
5. Paul Bauer
6. Prasun Bhattacharjee
7. James Booker
8. Fan Chen
9. Kai Chen
10. Michael Davis
11. Anupam Das
12. Richard Deitz
13. Brent Evans
14. Julie Anna Golebiewski
15. Prithviraj Guha
16. Rupayan Gupta
17. Tracy Henry
18. Michael Jerison
19. Robert Jones
20. Gisella Kagy
21. Eddery Lam
22. Joshua Mask
23. Michael McAvoy
24. Sean MacDonald
25. Alexander F. McQuoid
26. Albert Murphy
27. James Murtagh
28. Shahdad Nagshpour
29. William O’Dea
30. Michael O'Hara
31. Kasia Platt
32. Mitali Pradhan
33. Jungjun Park
34. Wei Simi
35. Philip Sirianni
36. Garrison Song
37. Jonathan Stone
38. Elias Shukralla
39. Della Lee Sue
40. Michael Toma
41. Masashi Tsutsumi
42. Pinar Uysal
43. Jeffrey Wagner
44. Ling Yan