2019

Attractive or Aggressive? A Face Recognition and Machine Learning Approach for Estimating Returns to Visual Appearance

Guodong Guo
West Virginia University, guodong.guo@mail.wvu.edu

Brad R. Humphreys
West Virginia University, brhumphreys@mail.wvu.edu

Mohammad I. Nouyed
West Virginia University, monouyed@mix.wvu.edu

Yang Zhou
West Virginia University, ygzhou@mix.wvu.edu

Attractive or Aggressive? A Face Recognition and Machine Learning Approach for Estimating Returns to Visual Appearance

Guodong Guo [*]
West Virginia University

Brad R. Humphreys [†]
West Virginia University

Mohammad Iqbal Nouyed[‡]
West Virginia University

Yang Zhou [§]
West Virginia University

August 16, 2019

Abstract

A growing literature documents the presence of appearance premia in labor markets. We analyze appearance premia in a high-profile, high-pay setting: head football coaches at big-time college sports programs. These employees face job tasks involving repeated interpersonal interaction on multiple fronts and also act as the "face" of their program. We estimate the attractiveness of each employee using a neural network approach, a pre-trained Convolutional Neural Network fine tuned for this application. This approach can eliminate biases induced by volunteer evaluators and limited numbers of photos. We also use this approach to estimate the perceived aggressiveness of each employee based on observable facial features. Aggressiveness can be detected from facial characteristics and may be a trait preferred by managers and customers in this market. Results show clear evidence of a salary premium for *less attractive* employees. No beauty premium exists in this market. We also find evidence of an aggressiveness premium, as well as evidence of higher attendance at games coached by less attractive and more aggressive appearing coaches, supporting customer based preferences for the premia. We also provide a methodological contribution by incorporating face recognition and computer vision analysis to evaluate employee appearance.

JEL Codes: C45, J71
Keywords: Beauty premium; facial recognition; machine learning; college football

[*]Benjamin M. Statler College of Engineering and Mineral Resources, Lane Department of Computer Science and Electrical Engineering, 109 Research Way, Morgantown, WV 26505, USA; Email: guodong.guo@mail.wvu.edu

[†]Corresponding author. John Chambers College of Business & Economics, Department of Economics, 1601 University Ave., PO Box 6025, Morgantown, WV 26506-6025, USA; Email: brhumphreys@mail.wvu.edu

[‡]Benjamin M. Statler College of Engineering and Mineral Resources, Lane Department of Computer Science and Electrical Engineering, 109 Research Way, Morgantown, WV 26505, USA; Email: monouyed@mix.wvu.edu

[§]John Chambers College of Business & Economics, Department of Economics, 1601 University Ave., PO Box 6025, Morgantown, WV 26506-6025, USA; Email: ygzhou@mix.wvu.edu

[1]We thank participants at the Eastern Economic Association Annual Conference and the Western Economic Association International Conference and seminar participants at the West Virginia University Department of Economics for helpful comments and suggestions.

1 Introduction

Labor market discrimination represents an important, widespread and widely studied economic outcome. Discrimination against unattractive workers recently became a prominent area in this literature. This outcome, called the "beauty premium," refers to the idea that more attractive people earn a premium in labor markets (Scholz and Sicinski, 2015). While the beauty premium may reflect employer or customer tastes, some research posits an alternative productivity-based explanation where physical attractiveness enhances worker productivity, primarily in jobs involving substantial interpersonal interaction (Stinebrickner et al., 2018).

The idea that attractiveness enhances productivity rests on heterogeneity in the beauty premium across job types, in terms of task requirements. Hamermesh et al. (1994) posited that the beauty premium affects people in jobs with substantial personal interaction, since attractiveness might be productivity enhancing in that setting. No productivity-enhancing effects of attractiveness should exist in jobs that involve working with numbers, information, or data.

The impact of attractiveness has been studied intensively in other social science disciplines like sociology and psychology. Hamermesh et al. (1994) undertook the first economic research on the topic. Since then, a growing body of economic research on the beauty premium emerged, including Biddle and Hamermesh (1998) analyzing lawyers' looks and lucre, Hamermesh and Parker (2005) analyzing instructors' pulchritude and student ratings, Mocan and Tekin (2010) analyzing physical attractiveness and criminal activity, Berggren et al. (2010) and Berggren et al. (2017) analyzing attractiveness and voter appeal, and others. A small, growing literature in sports economics also emerged, primarily focused on professional athletes (Berri et al., 2011; Ahn and Lee, 2014; Bakkenbüll and Kiefer, 2015; Dietl et al., 2018; Yamamura et al., 2018).

A lack of data sources containing both physical attractiveness measures and individual labor market outcomes represents an obstacle to economic research in this area. Hamermesh et al. (1994) exploited two novel large-scale surveys that contained a variable describing the attractiveness of each interview subject based on the opinion of the survey interviewer who observed the subject in person during the interview. Researchers often exploit existing one-off data that contain photographs and employ relatively small numbers of volunteer evaluators who examine the photos and rate the physical attractiveness of each or self assessed attractiveness (Mocan and Tekin, 2010). Many studies use relatively small numbers of evaluators; for example, Ahn and Lee (2014) use only 16 evaluators and Scholz and Sicinski (2015) use 12. Use of a small number of evaluators to generate attractiveness measures can introduce bias into the process because of evaluator preferences or measurement error associated with the photographs. This problem may be amplified by the common practice of rating a single photograph of individuals in the sample.[1] We provide a methodological contribution by developing a computer vision analysis of facial characteristics to assess attractiveness using standard tools from the computer science literature on facial recognition that can be readily applied to electronic images, eliminating the need for costly, time consuming

[1]Exceptions include (Scholz and Sicinski, 2015) who use five different photographs of each person in the sample, Dietl et al. (2018) (three photos) and Ahn and Lee (2014) (two photos).

human evaluation.

Research on the beauty premium among professional athletes faces an additional challenge: athletes evaluated in terms of attractiveness are famous and successful and evaluators may conflate success and fame with beauty, generating biased attractiveness measures. Berri et al. (2011) and Dietl et al. (2018) address this issue by using a one-dimensional computer vision measure of attractiveness, facial symmetry, to quantify attractiveness. Berri et al. (2011) use *Symmeter*[2] to rate the attractiveness of NFL quarterback's faces and generate a facial symmetry score for each. Dietl et al. (2018) use *Prettyscale*[3] to generate a facial symmetry score based on 14 facial landmarks manually placed on photos of 128 professional tennis players. Research in evolutionary biology (Perrett et al., 1999) supports the idea that facial symmetry represents an important one-dimensional component of attractiveness. We use a multidimensional approach for assessing attractiveness.

While economic research generally focuses on the impact of perceived attractiveness on economic outcomes, other research focuses on the impact of alternative observable facial characteristics. People frequently infer personal traits and characteristics based solely on observed facial features (Willis and Todorov, 2006) which could affect many outcomes. Todorov et al. (2005) investigate the impact of perceived candidate competence based on facial photographs on electoral outcomes and find that candidates perceived as more competent were more likely to win elections. Willis and Todorov (2006) assess the link between observed facial characteristics and five different individual traits (trustworthiness, competence, likeability, aggression, and attractiveness) and find that individuals form clear opinions about possession of each of these traits after very short exposure to photographs.

Mueller and Mazur (1996) found that West Point cadets perceived as having dominant faces, based on evaluators' assessments of yearbook photos, rose to higher ranks in the military over the course of their careers than cadets perceived as having less dominant facial characteristics. Duarte et al. (2012) found that individuals perceived as more trustworthy based on evaluations of photographs had higher credit scores and lower loan default rates. If perceived attractiveness generates labor market impacts, these other perceived personal characteristics based on photographs may also affect labor market outcomes, depending on the task requirements of specific jobs.

We investigate the effect of perceived employee facial characteristics in a high-pay occupation with tasks involving extensive personal interaction, including employee, peer, customer, and media interaction. The head coach of an National Collegiate Athletic Association (NCAA) football bowl subdivision (FBS) team, the largest and highest profile category of college football, interacts extensively with current players, recruits, assistant coaches, the media, fans, alumni, donors, referees, university administrators, and others. These interpersonal interactions have an important effect on job performance and may affect productivity. We find no evidence of a beauty premium in this setting, which provides important new evidence that limits to the beauty premium exist in jobs involving extensive interpersonal interactions. Instead, we find that less attractive head

[2] http://www.symmeter.com/

[3] http://www.prettyscale.com/

football coaches earn a salary premium relative to more attractive coaches after controlling for job performance, a novel finding in this literature.

Nearly all previous studies use volunteers, often students, to assess the attractiveness of individuals in the sample, often from a single photograph. We use a machine learning approach to identify specific facial features associated with attractiveness in a generic set of photographs and train a prediction model based on this evaluation. Our attractiveness measure comes from a machine learning approach that applies estimates from training on generic photographs to photographs of football coaches. This can mitigate some bias in attractiveness ratings due to the evaluation procedure. We also assess attractiveness based on multiple photographs for each employee, which can mitigate measurement error associated with single photographs of individuals.

We use 2,222 facial images and associated characteristics from the MIT 10k US Adult Faces Database (Bainbridge, 2017). We identify crucial facial features associated with men being attractive or unattractive and aggressive or not aggressive using a supervised learning approach. We use this information to rate the "attractiveness" and "aggressiveness" of all NCAA FBS head coaches from 2014 to 2016 based on publicly available facial photos and analyze the relationship between these facial features and salary, controlling for job performance. Empirical analysis reveals a beauty discount and aggressiveness premium in this labor market.

Another issue in the beauty premium literature is the possibility that attractive individuals may sort into jobs where they have a productivity advantage (jobs with substantial interpersonal interaction), which would generate a beauty premium in jobs with certain characteristics but would not reflect underlying discrimination against unattractive people. Our setting should not include sorting by attractiveness. Almost all head football coaches are former college and/or professional football players, which requires exceptional physical ability likely unrelated to attractiveness (nobody ever got a football scholarship because they were exceptionally attractive). So this specific occupation features a significant entry barrier that cannot be overcome by attractiveness alone, reducing the potential impact of sorting on attractiveness into occupations. Our finding of a premium for unattractive coaches may reflect widespread sorting in other occupations with no such entry barriers.

We also assess the impact of an alternative personal trait that can be determined from a photograph and could be either productivity-enhancing or preferred by employers and customers: aggressiveness. Little economic research analyzes the presence of earnings premia for personal traits other than attractiveness. Aggression represents an important characteristic of football players; this could carry over to coaches, given their goal of winning games and leading successful college football programs. We find evidence of a salary premium earned by coaches with facial characteristics perceived as aggressive. To our knowledge, no previous research identifies a salary premium associated with characteristics other than attractiveness in the economics literature.

Perceived aggressiveness based on observable facial characteristics could plausibly affect labor market outcomes. Aggressive behavior can be predicted from observable facial characteristics based on laboratory and field evidence (Carré and McCormick, 2008; Carré et al., 2009). A biological basis supports this relationship; research links an observable male facial characteristics, a larger facial width-to-height ratio, with higher measured levels of circulating blood testosterone (Lefevre et al., 2013). Increased blood testosterone could affect workplace productivity.

Finally, no evidence exists in the literature on the extent to which discrimination comes from employer-based preferences or customer-based preferences. We develop new evidence of fan based preferences for unattractive head coaches who look aggressive. More fans attend games coached by unattractive men relative to attendance at games coached by more attractive men, controlling for team success. This suggests that college football fans have a preference for less attractive head football coaches.

2 Beauty Advantage, Face Recognition, and Machine Learning

2.1 Beauty Advantage in Economics

Hamermesh et al. (1994) undertook the first research on the beauty premium in economics. This paper used interviewers' ratings of the physical appearance of respondents to three broad-based household surveys, two in the United States and one in Canada, to develop evidence supporting a beauty premium in the labor market. In these survey data sets, the interviewers rated the interviewee's physical attractiveness from 5 (homely) to 1 (strikingly attractive) during in-home interviews. Following this paper, most empirical research on the beauty premium used personal attractiveness measures based on the opinion of one or more individuals based on in-person observation of the individual being analyzed, or more commonly, a photograph.

Hamermesh et al. (1994) found that more beautiful people earned more, even in jobs where physical appearance had little to do with job performance. They reported a premium for above average attractiveness of about 5% and a penalty for below average attractiveness of about 7%. Their study suggests the existence of pure employer discrimination.

Following Hamermesh et al. (1994), a growing body of economics literature developed evidence of a beauty premium. Biddle and Hamermesh (1998) develop evidence of a beauty premium for lawyers. Hamermesh and Parker (2005) study the relationship between the teachers' beauty and find evidence suggesting that attractive teachers receive higher evaluations.

Mobius and Rosenblat (2006) further explored transmission channels from beauty to earnings in an experimental setting. 50 high school students rated individual attractiveness based on photographs of 330 experiment participants. They found a sizable beauty premium and identified three possible transmission channels: (1) more attractive people are more confident which increases their salaries; (2) even with equal confidence levels, employers consider more attractive people to have higher abilities; and (3) more attractive people actually have better "soft" skills like communication and social skills, which help them earn more.

Mocan and Tekin (2010) analyzed the relationship between attractiveness and criminal activity. They found that young adults with a more self-identified ugly look had a higher propensity to commit crimes, and that beauty in high school influenced teenagers' criminal behavior years later.

Berggren et al. (2010) investigated the beauty premium in political voting. If voters lack sufficient background information on candidates, then an overall impression of a candidate's physical appearance will influence voters' decisions. This happens frequently. Many voters only see photos of candidates on print media or images on other advertisements. This study used relatively large samples of both photos and reviews: 1,929 Finnish political candidates and evaluations by 10,011 respondents, of which 3,708 were Finnish. Berggren et al. (2017) further investigated whether the beauty premium differs for candidates with different political ideologies, because voters' preference may vary across social groups. They found that candidates on the right received higher attractiveness evaluations. One explanation for this is that beautiful people earn more, which makes them less inclined to support economic redistribution. This study also contained an experiment, which generated similar evidence.

Some existing research focuses on the existence of a beauty premium in sports. Sports represents an interesting setting for analyzing the beauty premium because of the ready availability of photos of athletes, salary data, and individual performance data. Berri et al. (2011) used the *Symmeter* program to assess the facial symmetry of photos of 138 NFL quarterbacks, along with annual performance and salary data over the period 1995 to 2009, and found evidence of a beauty premium. Facial symmetry represents one aspect of beauty. A change in facial symmetry from one standard deviation below the mean to one standard deviation above among NFL quarterbacks lead to an 11.8% salary increase, or about $338,000.

Ahn and Lee (2014) studied 132 female golfers appearing in at least one of the four major professional tournaments during the 2010 Ladies Professional Golf Association tour and annual performance and earnings data over the period 1992-2010. Ahn and Lee (2014) use ratings of photos from 16 volunteer evaluators as the beauty measure. Although they do not find any direct evidence of beauty premium leading more prize money in golf tournaments, they do find evidence that attractiveness generates an effort-enhancing effect: more beautiful golfers put more effort into the competition, perhaps because of a higher return to human capital for more attractive female golfers. This represents a potential mechanism through which beauty affects earnings. Ahn and Lee (2014) report that a one standard deviation increase in their beauty rating from the mean generated a $127,520 increase in tournament prize money, a 31% increase.

Yamamura et al. (2018) use direct evaluations of male and female Japanese speedboat racers' attractiveness to analyze the performance of these athletes. The attractiveness evaluations came from student recruits. They found that more attractive racers are more popular than their competitors, even after controlling conditions of the race, racer ability, and other characteristics. They further find that this "beauty premium" makes popular male racers perform better, but there is no such an effect for female racers. These results show that the productivity-enhancing effects of attractiveness in sport reported by Ahn and Lee (2014) generalize to sport, and may depend on gender.

Dietl et al. (2018) analyze the effect of attractiveness on television audience size in professional tennis using data from Switzerland. They analyze over 600 quarterfinal, semifinal and final Grand

Slam match broadcasts over the period 2000 to 2016 and use facial symmetry to proxy for attractiveness from the *Prettyscale* program. Unlike many other papers, they analyze the effect of attractiveness of both male and female tennis players. A one standard deviation in female facial symmetry increased television audience size by 24% but an increase in male facial symmetry did not increase television audience size. This suggests that customer discrimination could be the source of the beauty premium. The novel use of television audience size, a direct measure of demand, generates this.

Bakkenbüll and Kiefer (2015) analyzed the productivity-enhancing effect of attractiveness in female tennis players. Attractiveness measures come from an on-line survey administered to students who rated pictures of the athletes on an eight point scale. About 20 students rated each of 100 photos of female tennis players. They find that a one point increase in attractiveness increased annual prize money winnings by about 30% and career to date prize money winnings by about 20%, implying a sizable productivity effect for attractive female tennis pros.

All the research surveyed above used either direct ratings of photos from volunteer evaluators or one-dimensional geometric beauty measures (facial symmetry) from computer vision analysis of photos. The first approach may suffer from biases if using a small number of evaluators. The symmetry-based computer vision methods lag behind the current state of the art in facial recognition research. While facial symmetry probably matters, it may not be an adequate measure of attractiveness because it reflects only one aspect of facial characteristics.

Substantial research on facial recognition and the assessment of attractiveness exists beyond symmetry. Recent research in the computer science literature focuses on how to use more advanced computer-based methods to evaluate attractiveness and other characteristics that go beyond appearance from facial photos. Kagian et al. (2008) employed a machine learning predictor of facial attractiveness to reveal human-like psychological biases. Altwaijry and Belongie (2013) also used a machine learning approach to rate the attractiveness of photos. For more details of the current computer science literature on beauty rating, Laurentini and Bottino (2014) provide a very detailed overview.

Todorov et al. (2005) surveyed individuals who were asked to infer the competence of politicians from facial photos and then used these evaluations to predict voting results based on the assumption that voters prefer competent politicians. Difference in the photo-based evaluation of candidate competence predicted actual voting outcomes in 68% of the US Senate races in 2004, indicating that inferred personal traits based on facial photos provide useful information.

In terms of perceived aggressiveness based on facial characteristics, evidence of the importance of observable facial characteristics and actual aggressiveness exists. Prior research clearly links the perception of aggressiveness based on observable facial characteristics to actual aggressive behavior in the laboratory and the field in males. Carré and McCormick (2008) assessed the face width-to-height ratio (fWHR) of photographs of 88 undergraduate students and then performed laboratory assessments of their level of aggression. Males with a larger fWHR were more aggressive than those with a smaller ratio and also more aggressive than women with similar ratios. Carré et al.

(2009) report similar lab-generated results. Carré and McCormick (2008) also obtained photographs of every player on eight Canadian National Hockey League teams for the 2007-08 season and estimated their fWHR. Variation in the estimated face width-to-height ratio in this sample explained a significant portion of variation in aggressiveness, as measured by penalty minutes per game for each player over the course of the season. Players with larger fWHR spent more time in the penalty box per game than players with smaller fWHR.

A biological basis for a link between observable facial characteristics and aggression also exists. Observable facial characteristics, in particular the facial width-to-height ratio, has been linked to aggressive behavior in males and to higher testosterone levels. Lefevre et al. (2013) found that males with a higher fWHR had higher measured levels of circulating testosterone in their blood. Higher levels of testosterone causes more aggressive behavior in males.

3 Empirical Analysis

3.1 Data

The data used in this study come from three sources. The first consists of salary information for all NCAA FBS football head coaches from 2014 to 2016 taken from the USA Today college football coach database.[4] We augment the head coach salary data with annual team performance data from https://www.sports-reference.com/cfb/.

The second is the MIT US 10k Adult Faces Database (Bainbridge, 2017), which contains 10,168 facial photos reflecting the gender, age, and racial distributions of the adult U.S. population. From these 10,168 photos, 2,222 were randomly selected as target images and 6,468 constitute filler images. Information on 20 observable attributes were collected for each of the target images, including attractiveness and aggression using evaluations from two different Amazon Mechanical Turk (AMT) surveys with 12 and 30 different participants.

The final data source contains photos of all NCAA FBS football head coaches from 2014 to 2016 directly downloaded through Google Image Search. The method for downloading these photos included a search using the college and coach name as key words and then selecting the one most clear, large, and directly facing the camera.[5] Although we use our best judgment when selecting photos, some bias could be introduced through this process. Therefore, we download two additional photos for every coach and used the average rating of the three photos as the beauty measure.[6] The final sample contained 384 coach-year performance and salary observations from 2014 to 2016. The salary data contains 361 observations because of missing salary data for coaches employed at private universities, leaving 361 observations in the analysis sample.

[4] https://www.usatoday.com/sports/

[5] For consistency only one coauthor collected all photos.

[6] Most of the photos downloaded are facial. Some contain shoulders, arms, and other body parts. No photo contains just the face and nothing else, and even the facial photos contain at least some background images. Because the photos are "read" by computer a few photos with low pixels or other issues can not be processed. All three photos were read successfully for nearly all coaches.

Table 1: Summary Statistics

Statistic	Mean	St. Dev.	Min	Max
Wins	6.668	3.122	0	14
Losses	5.803	2.475	1	12
Salary ($000)	2,002	1,563	225	9,004
Age	50.817	8.202	34	77
Experience	9.269	8.018	0	34
Tenure at Current School	4.064	4.592	0	28
Attractiveness Score	3.227	0.569	1.403	4.519
Aggressiveness score	5.266	1.369	1.588	8.797

N=361

Table 1 contains summary statistics for all variables used in the empirical analysis. Win, the number of wins by a team in a season, varies from 0 to 14. $Loss$, the number of losses by a team in a season, varies from 1 to 12. A typical college football team plays about 12 games per season, although successful teams may play more. Depending on their performance in the regular season, a team attending playing in a bowl game or the Championship Playoffs might end up playing 14 games. $Salary$ is total salary of coach, which varies from $225,000 to more than 9 million US dollars. The empirical models use the log of this variable $Log\,Salary$ in the regression models.

We also capture the total years of head coaching experience ($Experience$) and tenure of each head coach at his current school, and their age. Note that $Experience$ and $Tenure\,at\,Current\,School$ are not the total years they have been a head coach or the head coach of this school, but are the number of years since the first time they served as a head coach or the head coach of this school. Data on coaches' age and career experience come from information on official team web sites and their Wikipedia pages.

The Attractiveness Score and Aggressiveness Score reflect evaluations of observable characteristics of each coach, based on the computer vision/machine learning approach discussed in detail in the following section. The machine learning approach used here generates continuous variables. These variables reflect the assessed attractiveness of each coach based on a continuous variable defined over the interval (1,5) where 5 is the most attractive coach and 1 is the least attractive, and the assessed aggressiveness of each coach based on a continuous variable defined over the interval (1,9) where 9 is the coach assessed as most aggressive and 1 is the least aggressive. The scales differ because the assessment questions appeared in different modules of an on-line survey. Most economics research uses discrete measures of attractiveness based on Likert-scale questions about attractiveness answered by volunteer evaluators.

Figure 1 contains selected photographs of coaches in the sample and their estimated attractiveness score to provide examples of unattractive and attractive head coaches. The attractiveness score for each photo is shown below the photo.

At the left of Figure 1, Tracy Claeys, head coach at the University of Minnesota in 2015 and

Figure 1: Attractiveness Scores for Selected Images

2016 has among the lowest attractiveness scores in the sample. Next, Kevin Sumlin, head coach at Texas A&M University throughout the sample, also has a relatively low attractiveness score. In the middle of Figure 1, Willie Taggert, head coach of the University of South Florida throughout the sample has an attractiveness score in the middle of the range, a slightly below average score. At the right on Figure 1, two of the coaches with higher attractiveness scores are Les Miles, head coach at LSU throughout the sample, and Tony Levine, head coach of the University Houston in 2014 has among the highest beauty scores in the sample.

Note that the photo of Les Miles has a different expression than the others. The machine learning approach used to generate beauty scores does not depend on facial expression. We know this because the pre-trained deep learning model used for this research has very high face recognition accuracy in real world face data sets where face images come from an unconstrained environment with varied expression, illumination, pose, and so forth (Wen et al., 2016).

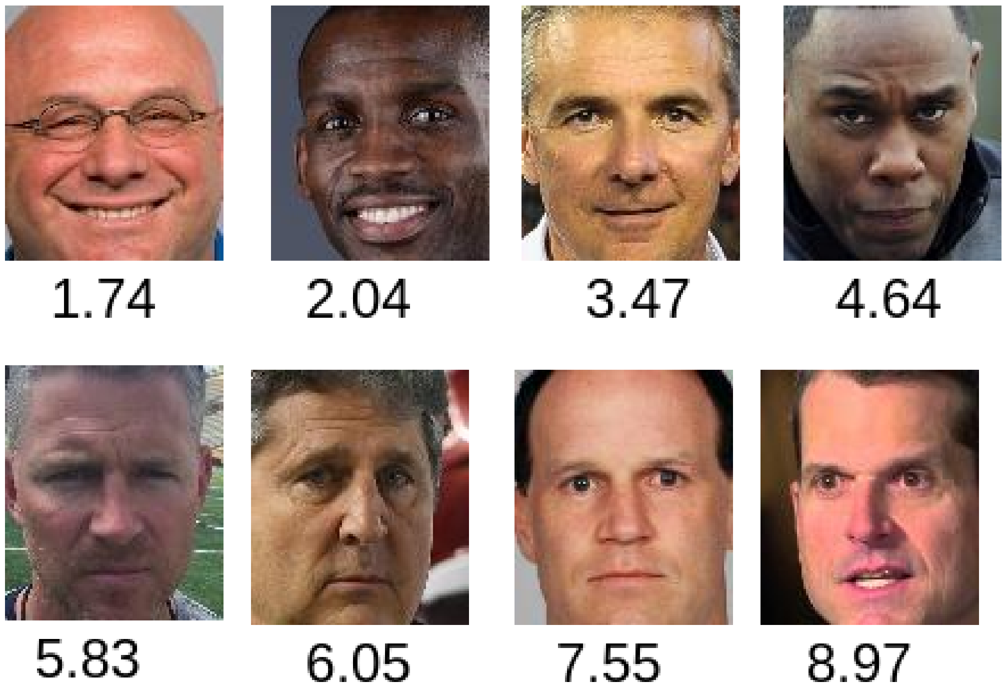
Figure 2 shows estimated aggressiveness scores for selected photographs. Recall that the aggressiveness score uses a 1 to 9 scale. At the top left of Figure 2, John Bonamego, head coach at Central Michigan University in 2015 and 2016 has among the lowest aggressiveness scores in the sample.

Ohio State coach Urban Meyer, third from the lest in the top row, has a below average aggression score (3.47). Venderbilt head coach Derek Mason (4.64) and Toledo head coach (in 2016) Jason Candle (5.83) have slightly below and slightly above average aggressiveness scores. At the bottom right of Figure 2, Jim Harbaugh, Michigan head coach in 2015 and 2016 has among the highest aggressiveness scores in the sample. Note that attractiveness scores and aggressiveness scores are correlated (see Figure 6 for details).

3.2 Face Recognition and Machine Learning

The facial characteristic evaluation scores for employees in the sample come from a computer vision machine learning approach. This section describes the neural network architecture used for training the beauty and aggressiveness prediction models and provides details about the training set construction.

Figure 2: Aggressiveness Scores for Selected Images



We begin from a standard source of facial photographs, the MIT 10k US Adult Face Database.[7] This database of facial images contains about 10,186 natural photographic images. The photos are representative of the US population in terms of age, gender, and race. 2,222 of the photos have a rich set of attributes attached to them, including attractiveness ratings with a range from 1 to 5 (a higher score means more attractive) and aggressiveness ratings with a range of 1-9. See Bainbridge et al. (2013) for details on the database and Khosla et al. (2013) for details on the attributes.

The next section contains details on how the attractiveness and aggressiveness scores were estimated. The procedure uses a Convolutional Neural Network, a standard neural network architecture.

3.2.1 Convolutional Neural Network (CNN) Architecture

The face image database and corresponding facial attributes data, can be used to train a neural network to analyze facial characteristics of any photograph. We use a pre-trained Convolutional Neural Network (CNN) architecture model proposed by Wen et al. (2016) for face recognition, and perform a transfer learning approach called "fine-tuning" to learn to predict attractiveness and aggressiveness scores for our assessment of facial characteristics of FBS head football coaches. Again, the use of pre-trained CNNs and transfer learning for the analysis of facial photos is standard in the machine learning. This CNN has been used extensively in the literature (Wen et al., 2016).

CNNs use filters to extract image characteristics. A filter is a 2-dimensional array applied to the

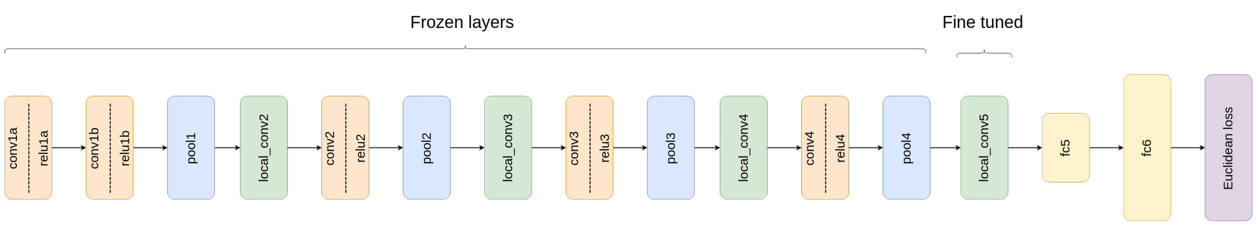
[7] https://www.wilmabainbridge.com/facememorability2.html.

digitized face image to extract specific features. Each filter element contains a set of weights, and these weights are updated iteratively based on a loss function using a back-propagation algorithm. The gradient descent optimization algorithm uses back-propagation to ensure that the weights are updated in such a way that the loss decreases as training images are iteratively passed through the network. At each training iteration, the weights are updated for the entire training set as the value of the loss function decreases. After a certain number of iterations, if the value of the loss function is reasonably small, the network has successfully learned to discriminate among face images with different attractiveness (or aggressiveness) levels.

A deep CNN is a hierarchical feature learner. At the bottom layers it learns to discriminate among small, finer aspects of photos, such as edges, textures etc, and in the subsequent layers progressively larger image characteristics are analyzed, such as parts of faces or whole face area.

The fine tuning, or transfer learning, process freezes all convolution layers except the local_conv5 layer, as shown in Figure 3. The local convolution layer is a group of convolutions pooling and rectifying linear unit operations, followed by an element wise operation to combine the output with the pooling layer. The freezing retains the weights for low level features learned by training using the original large-scale face data set while tuning the higher level features capturing characteristics reflecting beauty or aggression. Figure 3 shows the CNN architecture and fine tuning components of the process. Note that, even though the same architecture is used, separate neural networks were trained to perform the transfer learning for attractiveness and aggressiveness score prediction.

Figure 3: CNN Architecture (Wen et al., 2016) Frozen and Fine Tuning Layers



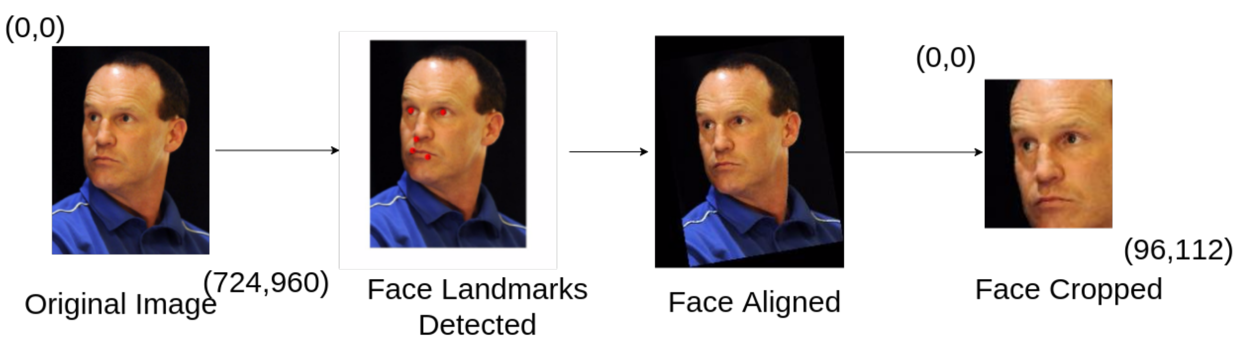
While the training and test photo sets come from different sources, both sets of photos contain images with good resolution and close frontal poses, well illuminated faces, and sharp images, which are the major sources of quality degradation in computer vision. We assume both the training and test photo sets have the same distribution of facial characteristics. Using a separate beauty dataset, SCUT-FBP5500 (Liang et al., 2018), and the associated different training protocol, we used 10 different common deep learning face recognition models with alternative parameter tuning and layer freezing approaches. We found the Wen et al. (2016) deep architecture to produce the best predictive results in terms of mean absolute error (MAE) and mean squared error (MSE) among this group of models. We use the same architecture for attractiveness and aggressiveness prediction because, to the best of our knowledge, this is the first time facial aggressiveness prediction has been done using a deep learning model. We lack an earlier source as a reference to select a suitable training model. Since this architecture performed reasonably well for predicting attractiveness, it should provide reasonable estimates of aggressiveness after fine tuning.

3.2.2 CNN Training Details

Preprocessing

Initially, facial landmarks on photographs of football coaches were identified using the method proposed by Baltrusaitis et al. (2013). This method identifies 68 separate facial landmark points like eyebrows, eyes and lips. We use 5 of these landmarks in preprocessing: Eye centers, nose tip, and mouth corners. Using these landmarks, we align each face image using an affine transformation, crop the image to a uniform size (96×112) and save it in RGB format. Figure 4 shows the face landmark detection, alignment and cropping phases for a sample face image from the football coach photos analyzed. The photo is of Air Force Academy head coach Troy Calhoun.

Figure 4: Face Preprocessing - Sample Image



Training Data

We collect 2,222 images with information on attractiveness and aggression from the MIT data base and use a data augmentation process to increase the size of the training set to improve the neural network performance. Since we quantify beauty and aggression, we need to be careful not to distort images in a way that loses finer image features that convey information about attractiveness or aggression. We avoided adding noise, smoothing, or adjusting the contrast and brightness of the images to create new samples for the training set. Instead, new samples were created using mirroring, rotation, and shift. For each sample, 39 additional images were created through this process for each existing photo in the training set, which leads to an augmented image data set with $2{,}222 \times 40 = 88{,}880$ observations. Training labels for facial attractiveness range from 1 to 5 with continuous values, where 1 indicates least attractive and 5 most attractive. For aggression, the ratings range from 1 to 9 with 1 indicating least aggressive and 9 the most aggressive. Note that, discrete score levels from the two AMT surveys were used to assess attractiveness and aggressiveness, the ratings for each image from the human raters were averaged to generate a single score per image.

This real valued score is used as the training label for the face images.

Convolutional Neural Network (CNN) Settings

To update the weights on the labels we use a Euclidean loss function. Because we work with training labels with continuous values this can be addressed as a regression problem. The Euclidean loss function used to update the weights on the labels is defined as:

$$E = \frac{1}{2N} \sum_{n=1}^{N} ||\hat{y}_n - y_n||_2^2 \tag{1}$$

where N is the number of training samples, y_n is the ground truth label from the MIT 10k database, and $\hat{y}_n$ is the predicted label from the CNN. The training was done over 300,000 iterations, using a batch size of 10. The training used a fixed learning rate $1e{-7}$, momentum 0.9, and weight decay 0.0001.

After training, the football coach photos are analyzed by the corresponding trained CNN to generate attractiveness and aggression scores for each coach's photo. These scores are averaged across photos for each coach to arrive at the attractiveness and aggressiveness values used in the empirical analysis. The attractiveness score for every coach is continuous over the interval 1-5 and the aggressiveness score is continuous over the interval 1-9. The intervals differ in the original MIT 10k data. From Table 1, the minimum average attractiveness score of a coach is 1.403, the maximum is 4.519, and the mean is 3.227. The aggressiveness score variable has substantially more variation around the mean than the attractiveness score variable.

3.3 Econometric Methods

We use an OLS model to explain observed variation in annual salaries earned by NCAA FBS head coaches. The key explanatory variables of interest are measures of the perceived attractiveness and aggressiveness of each coach based on a machine learning approach to evaluating their photographs. Like Berri et al. (2011), Dietl et al. (2018), Stinebrickner et al. (2018) and others in this literature, we use a single attractiveness and aggressiveness score value for each coach over the sample and estimate a pooled OLS model, even though we explain variation in salaries over three seasons. Like Berri et al. (2011), we assess the validity of this approach by estimating alternative model specifications that collapse all variables to sample averages.

The main empirical model is

$$S_{ist} = \beta A_i + \psi Z_{ist} + \epsilon_{ist}. \tag{2}$$

S_{ist} denotes the (log) salary of college football coach i coaching at school s in season t. A_i is the attractiveness score or alternatively the aggressiveness score estimated for coach i. Z_{ist} includes control variables reflecting coach and team performance and characteristics: number of wins by the football team under coach i at school s in season t, total years of experience for coach i in season t

and coach i's tenure at school s. ϵ_{ist} is the error term. The error term is assumed to be distributed with mean zero and constant variance.

Note that, like all papers in this literature, we cannot identify a separate head coach fixed effect and an effect of attractiveness and aggression. We can only identify and estimate a school fixed effect, based on schools who employed more than one coach in the sample. 40 of the 128 schools in the sample changed coaches, including three at Illinois, so we should adequately estimate school fixed effects.

4 Beauty and Aggressiveness Premia Estimates

4.1 Nonparametric Graphic Results

We first generate nonparametric graphic results to show the basic relationship between attractiveness score and salaries in the data. We first estimate a model explaining log salary of coach i at school s in year t

$$Sist = \psi Z_{ist} + \epsilon_{ist}. \tag{3}$$

using coach/school characteristics and school and season fixed effects. We then take the residuals from this model and estimate a local polynomial regression model that uses higher-order functions of the attractiveness score and separately the aggressiveness score to generate nonparametric salary gradients as functions of attractiveness and aggressiveness. The 95% confidence interval for the local polynomial salary gradients are shown in Figure 5.

Figure 5 shows the basic attractiveness (left panel) and aggressiveness (right panel) results. Controlling for coach success, age, and experience, less attractive coaches earn higher salaries. The effect becomes significant at attractiveness scores a bit over 2. Some nonlinearities exist, but attractive head football coaches do not earn a premium. Instead, unattractive coaches earn a premium relative to attractive coaches.

The salary gradient for the aggressiveness score shows evidence of an aggressiveness premium for aggressiveness scores over about 8. This relationship might exhibit nonlinearity that takes a U shape, although the significance of the salary increase at the low end of the aggressiveness score distribution appears weak. Head coaches assessed as very aggressive earn premium, while those in the middle of the distribution do not.

4.2 Pooled OLS Model Results

The primary regression results are displayed in Table 2. Given the evidence from the salary gradients above, we include squared terms for the attractiveness and aggressiveness scores to capture any nonlinearities in the relationships. These models control for coach productivity and experience, but do not control for school or year fixed effects.

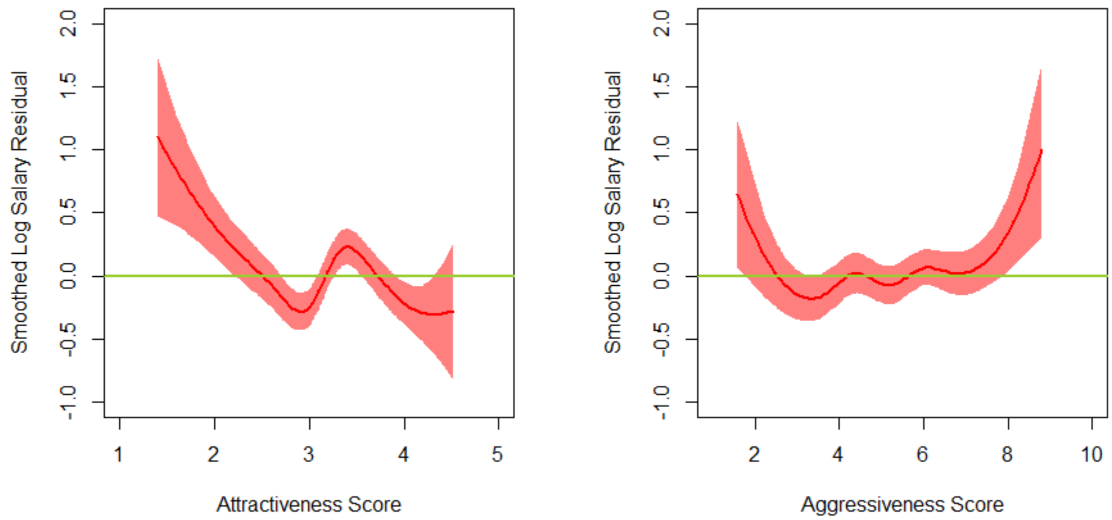
More successful or productive coaches earn higher salaries across all model specifications. The return to experience at the coach's current school, his tenure at that school, is also positive across

Table 2: Beauty and Aggressiveness Premia - Pooled OLS Model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Attractiveness	-0.222^{***}	-0.965^{*}			-0.315^{***}	-1.433^{**}
	(0.077)	(0.529)			(0.082)	(0.576)
Attractiveness^2		0.120				0.179^{**}
		(0.085)				(0.090)
Aggressiveness			0.056^{*}	-0.245	0.103^{***}	0.025
			(0.032)	(0.168)	(0.033)	(0.181)
Aggressiveness^2				0.029^{*}		0.009
				(0.016)		(0.017)
Wins	0.088^{***}	0.090^{***}	0.085^{***}	0.087^{***}	0.089^{***}	0.093^{***}
	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)
Log Age	-0.191	-0.199	0.402	0.322	0.008	0.016
	(0.402)	(0.402)	(0.397)	(0.398)	(0.403)	(0.402)
Experience	-0.000	0.001	-0.005	-0.003	-0.002	0.000
	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Tenure at Current School	0.049^{***}	0.050^{***}	0.044^{***}	0.046^{***}	0.051^{***}	0.052^{***}
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
R^2	0.199	0.204	0.188	0.195	0.220	0.232
N	361	361	361	361	361	361

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Figure 5: Salary Gradients - Local Polynomial Model



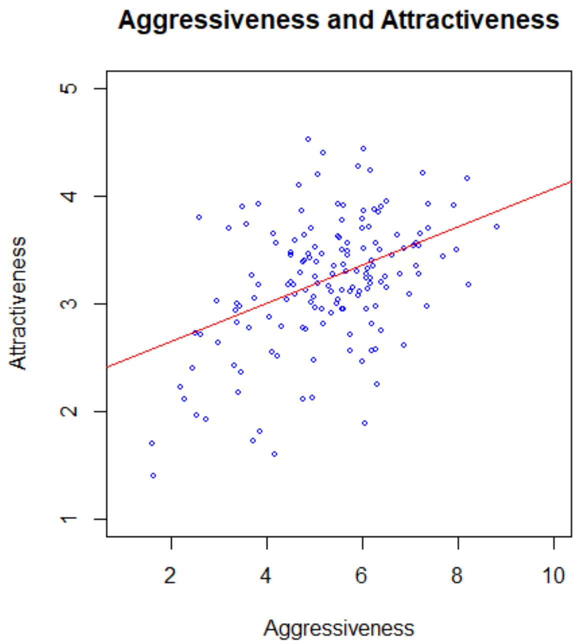
all model specifications.

Model 1 contains only the attractiveness score and Model 2 adds the square of the attractiveness score. The premium paid to unattractive head coaches clearly emerges in these results, although the results are stronger when the attractiveness score enters linearly. The salary premium for a one unit decrease in the attractiveness score is 0.222 log points, or a salary increase of about 22%. A one unit decrease is a bit less than two standard deviations for the attractiveness score variable, based on the reported standard deviation on Table 1.

Model 3 contains only the aggressiveness score and Model 4 adds the square of the aggressiveness score. Evidence of an aggressiveness premium is present but weak in Models 3 and 4. While a 10% significance level may seem generous, the data set contains only 361 observations. One reason for the weaker relationship may be the U-shape of the salary gradient for aggressiveness shown on Figure 3. The estimated aggressiveness premium for Model 3 is equivalent to a 5.6% increase in salary for a one unit increase in the aggressiveness score. A one unit increase in the aggressiveness score is a change of less than one standard deviation.

Model 5 includes both the aggressiveness score and the attractiveness score, generating attractiveness premium estimates holding aggressiveness constant and aggressiveness premium estimates holding attractiveness constant. Figure 6 shows the relationship between aggressiveness and attractiveness in the sample. Given the positive relationship, both variables belong in the same regression model to avoid omitted variables bias.

Figure 6: Aggressiveness and Attractiveness



$\rho_{pearson} = 0.426$, $\rho_{kendall} = 0.241$, $\rho_{spearman} = 0.346$

Model 5 results contain clear evidence of both an aggressiveness premium holding attractiveness constant and a premium for unattractive head coaches holding aggressiveness constant. The unattractiveness premium for a one point decrease in the attractiveness score is about 32% and the aggressiveness premium is about 10% for a one point increase in the aggressiveness score.

4.3 Fixed Effects Model Results

A first set of robustness checks are shown in Table 3. These models control for coach productivity and experience, and both school and year fixed effects. These models control for any potential correlation between unobservable time-invariant school-specific effects and the attractiveness and aggressiveness of the school's head football coach. For example, if administrators at a particular school tended to hire particularly attractive or aggressive head football coaches.

$$S_{ist} = \beta A_i + \psi Z_{ist} + C_s + Y_t + \epsilon_{ist}. \tag{4}$$

Where every item in Equation (4) is the same as that in Equation 2, but two fixed effects are added. C_s is a school fixed effect and Y_t is a season fixed effect.

First, note that including school fixed effects in the empirical model changes the statistical significance of the estimated parameter on the wins variable. This three season period contains

Table 3: Beauty and Aggressiveness Premia - School Fixed Effects Model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Attractiveness	-0.122^{**}	0.381			-0.158^{**}	0.671^{*}
	(0.051)	(0.316)			(0.062)	(0.346)
Attractiveness^2		-0.085				-0.134^{**}
		(0.053)				(0.056)
Aggressiveness			-0.012	-0.297^{***}	0.028	-0.339^{***}
			(0.023)	(0.098)	(0.027)	(0.105)
Aggressiveness^2				0.028^{***}		0.034^{***}
				(0.009)		(0.010)
Wins	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Log Age	0.416^{*}	0.436^{*}	0.591^{***}	0.637^{***}	0.425^{*}	0.517^{**}
	(0.230)	(0.230)	(0.223)	(0.220)	(0.230)	(0.225)
Experience	-0.002	-0.003	-0.005	-0.006	-0.001	-0.005
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Tenure at Current School	-0.006	-0.005	-0.006	-0.004	-0.005	-0.003
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
School FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	361	361	361	361	361	361
R^2	0.976	0.976	0.975	0.976	0.976	0.977

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

relatively low variation in the number of wins across the sample, reflecting a general lack of churn in terms of successful and unsuccessful teams. From Table 1 the standard deviation of the wins variable is about 3 wins per season. If successful teams win between 6 and 12 games per season (9 ± 3) mediocre teams between 3 and 9 games (6 ± 3) and unsuccessful teams between 0 and 6 games (3 ± 3) on average, then the addition of a school fixed effect would be expected to reduce the estimated statistical significance of the wins variable. Also, the estimated parameter on the age variable becomes statistically different from zero with the predicted sign.

Evidence of beauty and aggressiveness premia remain on Table 3. Again, Models 1 and 2 contain only the attractiveness score and its square, Models 3 and 4 contain only the aggressiveness score and its square, and Models 5 and 6 contain both. The results for Model 1 support a linear attractiveness discount. Results for Model 4 support a nonlinear, U-shaped aggressiveness premium. Taken together, Results from Models 5 and 6 support the presence of both. The main results appear robust to the inclusion of school and year fixed effects.

4.4 Collapsed Average Model Results

Again, the attractiveness and aggressiveness scores effectively represent a coach fixed effect. We follow Berri et al. (2011) and check the robustness of the main results by estimating alternative models that collapse the sample to average values. This reduces the sample size by a factor of three, so some loss of statistical significance should be expected. Figure 7 shows the nonparametric salary gradients for the attractiveness and aggressiveness scores. The salary gradient for the attractiveness score is virtually unchanged from the full sample results on Figure 5. The salary gradient for the aggressiveness score loses significance at the ends of the distribution, but the general U-shaped relationship still appears.

Table 4 contains the corresponding regression results. The layout of the results on Table 4 are the same as for Tables 2 and 3 above. Models 1 and 2 contain only the attractiveness score and its square, Models 3 and 4 contain only the aggressiveness score and its square, and Models 5 and 6 contain results with both variables included.

In general, the results on Table 4 support the robustness of the main results, a premium paid to unattractive coaches and to coaches who appear to be aggressive. While many estimated parameters lose statistical significance due to the reduced sample size, the estimated parameters on the attractiveness and aggressiveness score variables in Model 5 strongly resemble those from Model 5 from the pooled OLS results on Table 2.

5 Fan Preferences

5.1 Fan Preferences Results with Attendance Data

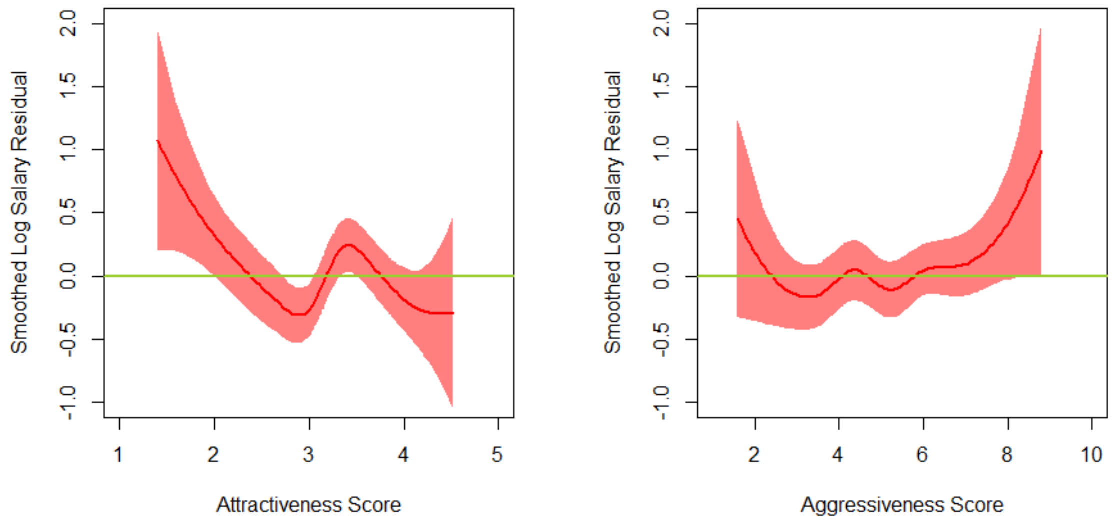
Why do unattractive and aggressive NCAA FBS head football coaches earn salary premia? Theory identifies two potential sources: discrimination on the part of either managers or fans, or alternatively the personality traits associated with perceived facial attractiveness or aggression increase

Table 4: Beauty and Aggressiveness Premia - Pooled OLS Collapsed Average Model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Attractiveness	-0.182^{*}	-1.032			-0.301^{**}	-1.523^{*}
	(0.109)	(0.744)			(0.117)	(0.793)
Attractiveness^2		0.140				0.200
		(0.121)				(0.126)
Aggressiveness			0.069	-0.266	0.120^{**}	-0.011
			(0.045)	(0.240)	(0.048)	(0.254)
Aggressiveness^2				0.033		0.015
				(0.023)		(0.024)
Wins	0.131^{***}	0.135^{***}	0.126^{***}	0.129^{***}	0.134^{***}	0.142^{***}
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)
Log Age	0.072	0.075	0.619	0.542	0.284	0.301
	(0.583)	(0.582)	(0.575)	(0.576)	(0.579)	(0.578)
Experience	0.000	0.002	-0.005	-0.003	-0.002	0.001
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Tenure at Current School	0.035^{**}	0.035^{**}	0.031^{*}	0.033^{*}	0.037^{**}	0.038^{**}
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
R^2	0.229	0.235	0.226	0.237	0.259	0.277
N	156	156	156	156	156	156

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Figure 7: Collapsed Sample Salary Gradients - Local Polynomial Model



productivity. The empirical models above control for head coach productivity by including the number of wins for each team in the model. The school fixed effects models should partially control for preferences of administrators for attractive or aggressive head football coaches, to the extent that these preferences are time-invariant over the three year sample period.

In this context, we have access to information on fan preferences for head coach attractiveness or aggressiveness in the form of attendance at football games. If fans prefer to watch college football games coached by attractive or aggressive, then attendance at games coached by attractive or aggressive coaches would be higher, other things equal.

To investigate the role played by fan preferences in driving the attractiveness and aggressiveness premia documented above, we replace the log salary dependent variable in Equation (2) with the log of total attendance at home games played by the team at school s coached by coach i in season t and re-estimate the model. Table 5 lists the summary statistics for the observations when using home average attendances as the dependent variables. The data source for home average attendances are from the official website of NCAA.[8] Since the data for home attendance have have full information, we have 380 observations in the analysis.

Table 6 contains results for the pooled OLS model, and the layout of the results on Table 6 are

[8]2014, http://fs.ncaa.org/Docs/stats/football_records/Attendance/2014.pdf. 2015, http://fs.ncaa.org/Docs/stats/football_records/Attendance/2015.pdf. 2016, http://fs.ncaa.org/Docs/stats/football_records/Attendance/2016.pdf.

Table 5: Summary Statistics for Attendance Regressions

Statistic	Mean	St. Dev.	Min	Max
Wins	6.639	3.101	0	14
Losses	5.821	2.463	1	12
Average Home Attendance	42,424	25,985	4,897	110,468
Age	50.639	8.162	34	77
Experience	9.068	8.015	0	34
Tenure at Current School	3.903	4.562	0	28
Attractiveness Score	3.221	0.571	1.403	4.519
Aggressiveness score	5.260	1.352	1.588	8.797

N=380

Table 6: Beauty and Aggressiveness Scores and Home Attendance - Pooled OLS Model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Attractiveness	-0.102^{*}	-0.780^{**}			-0.160^{***}	-1.106^{***}
	(0.054)	(0.367)			(0.058)	(0.395)
Attractiveness^2		0.110^{*}				0.153^{**}
		(0.059)				(0.062)
Aggressiveness			0.043^{*}	-0.147	0.066^{***}	0.023
			(0.023)	(0.120)	(0.024)	(0.128)
Aggressiveness^2				0.019		0.006
				(0.012)		(0.012)
Wins	0.085^{***}	0.087^{***}	0.082^{***}	0.084^{***}	0.086^{***}	0.089^{***}
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Log Age	0.189	0.201	0.507^{*}	0.457	0.334	0.375
	(0.282)	(0.281)	(0.280)	(0.281)	(0.285)	(0.284)
Experience	-0.005	-0.003	-0.007	-0.006	-0.006	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Tenure at Current School	0.020^{**}	0.021^{***}	0.018^{**}	0.019^{**}	0.021^{***}	0.022^{***}
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.211	0.219	0.212	0.217	0.227	0.242
N	380	380	380	380	380	380

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

the same as the other tables above. Models 1 and 2 contain only the attractiveness score and its square, Models 3 and 4 contain only the aggressiveness score and its square, and Models 5 and 6 contain results with both variables included.

The results strongly resemble those using log salary in terms of sign and statistical significance of parameter estimated. Winning teams draw more fans, other things equal. The longer the tenure of the current coach at a school, the higher is attendance, other things equal.

For the parameters of interest that reflect the effect of the facial attractiveness and aggressiveness of the head coach, Model 5 again generates results indicating that more fans attend games coached by less attractive head coaches, and more fans attend games coached by men who's facial characteristics are more aggressive. Since revenues increase with attendance, these results indicate that fan preference for head football coaches with specific facial appearances drives some of the attractiveness and aggressiveness premia present in this market.

Figure 8 shows the nonparametric attendance gradient based on estimation of Equation (3) using log home attendance as the dependent variable. Again, this model contains coach success, age, and experience and regresses the residuals from Equation (3) on higher order polynomial functions of the attractiveness score and the aggressiveness score for each head coach.

Figure 8: Attendance Gradients - Local Polynomial Model

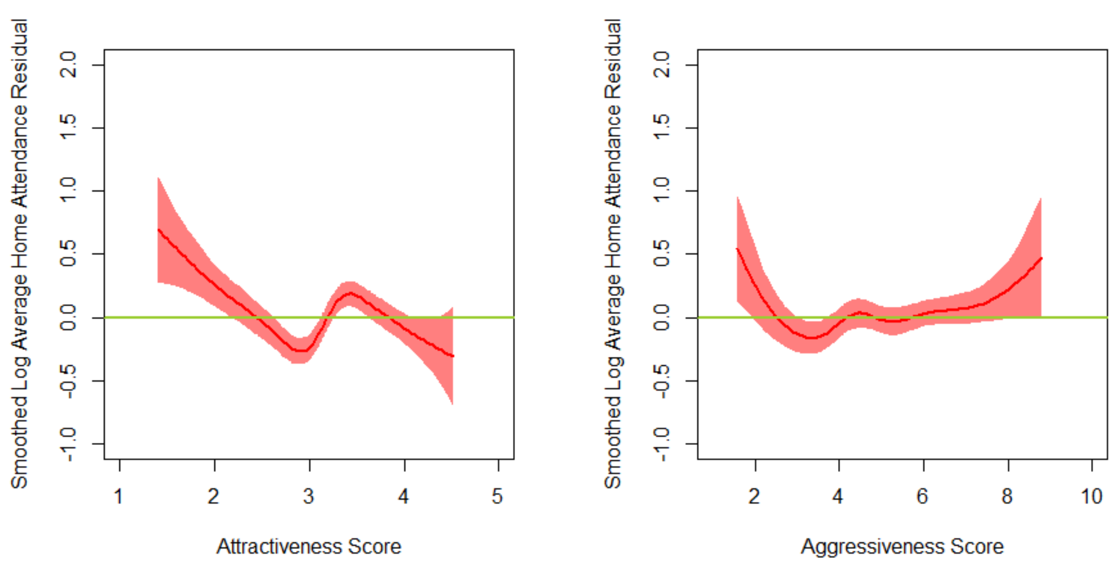


Figure 8 also shows clear evidence of higher attendance at games coached by less attractive men, holding team success, coach age, experience, and tenure. College football fans exhibit preferences for less attractive head coaches. The nonlinear U-shape in the attendance gradient for aggressive appearance can also be seen, but the statistical significance of this effect is weaker.

5.2 Mechanisms for a "Reverse" Beauty Premium

Why do less attractive FBS head coaches earn a salary premium compared to more attractive coaches? The existing literature generally finds a beauty premium for men but not for women. Attractiveness generates a differential salary effect across genders. The underlying mechanism could be that the effect of attractiveness on manager or customer decisions also varies by gender. For example, attractive male lawyers may have a larger effect on decisions made by female jurors than on male jurors. Attractive male salesmen may have a bigger influence on female customers than attractive female saleswomen on male customers.

One important form of personal interaction in this occupation is to convince 17 to 18 year old football recruits to play on their team. This is primarily male-male personal interaction. In this type of personal interaction, being a less attractive male might be productivity enhancing, as a less attractive male recruiter could be less threatening or intimidating to a potential recruit. More attractive male recruiters could raise issues related to sexual orientation that adolescent males might find uncomfortable. Little known about heterogeneity in the effect of beauty on managers or customers. In terms of attendance, most college football fans are male and might have a similar aversion to attractive male coaches ("I'm not watching that pretty-boy coach my team.") This could generate lower attendance.

Another explanation is that some other aspects of unattractiveness may be productivity enhancing in college football coaching. For example, if unattractive men are less likely to be married, then they would have more time to spend on the long hours involved in coaching football at the highest level, especially early in their career.

6 Conclusions

In this paper, we study the beauty premium of college football coaches in United States from 2014 to 2016. Using a face recognition and machine learning approach, we digitize and map 2,222 photos from the MIT Face Photo Database and capture the crucial features of this mapping. With the beauty reviews combined, we capture the main features and their contributions to reviews' evaluation of the beauty degree of a face photo. We further digitize the face photos of all the U.S. college football coaches from 2014 and 2016 and evaluate their beauty based on the surprised learning approach.

Empirical analysis suggests that a more attractive football coaches actually face a salary discount, rather than a premium. Instead, FBS head coaches with a more aggressive visage earn a salary premium. One explanation for an attractiveness discount and aggressiveness premium may stem from the fact that American football is a very aggressive sport, and an unattractive face might signal mental and physical toughness, viewed as a desirable characteristic in this market. We find evidence suggesting an overall premium for more aggressive coaches, and also evidence of a non-linear U-shaped relationship. Various robustness tests all suggest that our findings of a beauty discount and aggressiveness premium are robust.

Our research contributes to the current literature in three ways. First, the face recognition and machine learning approach is new and straightforward to implement, which enables researchers to evaluate the attractiveness of any facial photo with sufficient pixels, rather than having to recruit evaluators to sit down and do manual assessments. This comprehensive face mapping and machine learning approach reflects a more complex and comprehensive measure of attractiveness than one-dimensional measures like symmetry. Second, the beauty discount found contracts results in the literature, which suggests that whether commonly perceived beauty generates a premium or discount reflects specific labor market factors. Finally, our paper is the first to study and find evidence of an aggressiveness premium, which extend economists' understanding of the extent of observable factors influencing labor market outcomes.

References

Ahn, S. C. and Lee, Y. H. (2014). Beauty and productivity: The case of the Ladies Professional Golf Association. *Contemporary Economic Policy*, 32(1):155–168.

Altwaijry, H. and Belongie, S. (2013). Relative ranking of facial attractiveness. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 117–124. IEEE.

Bainbridge, W. A. (2017). The memorability of people: Intrinsic memorability across transformations of a person's face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(5):706.

Bainbridge, W. A., Isola, P., and Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, 142(4):1323.

Bakkenbüll, L.-B. and Kiefer, S. (2015). Are attractive female tennis players more successful? an empirical analysis. *Kyklos*, 68(4):443–458.

Baltrusaitis, T., Robinson, P., and Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 354–361.

Berggren, N., Jordahl, H., and Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, 94(1-2):8–15.

Berggren, N., Jordahl, H., and Poutvaara, P. (2017). The right look: Conservative politicians look better and voters reward it. *Journal of Public Economics*, 146:79–86.

Berri, D. J., Simmons, R., Van Gilder, J., and O'Neill, L. (2011). What does it mean to find the face of the franchise? physical attractiveness and the evaluation of athletic performance. *Economics Letters*, 111(3):200–202.

Biddle, J. E. and Hamermesh, D. S. (1998). Beauty, productivity, and discrimination: Lawyers' looks and lucre. *Journal of Labor Economics*, 16(1):172–201.

Carré, J. M. and McCormick, C. M. (2008). In your face: facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings of the Royal Society B: Biological Sciences*, 275(1651):2651–2656.

Carré, J. M., McCormick, C. M., and Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, 20(10):1194–1198.

Dietl, H., Özdemir, A., and Rendall, A. (2018). The role of physical attractiveness in tennis tv-viewership. Technical Report 376, University of Zurich, Department of Business Administration (IBW).

Duarte, J., Siegel, S., and Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8):2455–2484.

Hamermesh, D. S., Biddle, J. E., et al. (1994). Beauty and the labor market. *American Economic Review*, 84(5):1174–1194.

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

Kagian, A., Dror, G., Leyvand, T., Meilijson, I., Cohen-Or, D., and Ruppin, E. (2008). A machine learning predictor of facial attractiveness revealing human-like psychophysical biases. *Vision Research*, 48(2):235–243.

Khosla, A., Bainbridge, W. A., Torralba, A., and Oliva, A. (2013). Modifying the memorability of face photographs. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3200–3207.

Laurentini, A. and Bottino, A. (2014). Computer analysis of face beauty: A survey. *Computer Vision and Image Understanding*, 125:184–199.

Lefevre, C. E., Lewis, G. J., Perrett, D. I., and Penke, L. (2013). Telling facial metrics: facial width is associated with testosterone levels in men. *Evolution and Human Behavior*, 34(4):273–279.

Liang, L., Lin, L., Jin, L., Xie, D., and Li, M. (2018). SCUT-FBP5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1598–1603. IEEE.

Mobius, M. M. and Rosenblat, T. S. (2006). Why beauty matters. *American Economic Review*, 96(1):222–235.

Mocan, N. and Tekin, E. (2010). Ugly criminals. *The Review of Economics and Statistics*, 92(1):15–30.

Mueller, U. and Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social Forces*, 74(3):823–850.

Perrett, D. I., Burt, D. M., Penton-Voak, I. S., Lee, K. J., Rowland, D. A., and Edwards, R. (1999). Symmetry and human facial attractiveness. *Evolution and Human Behavior*, 20(5):295–307.

Scholz, J. K. and Sicinski, K. (2015). Facial attractiveness and lifetime earnings: Evidence from a cohort study. *The Review of Economics and Statistics*, 97(1):14–28.

Stinebrickner, T., Stinebrickner, R., and Sullivan, P. (2018). Beauty, job tasks, and wages: A new conclusion about employer taste-based discrimination. *The Review of Economics and Statistics*, (in press).

Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626.

Wen, Y., Zhang, K., Li, Z., and Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016*, pages 499–515, Cham. Springer International Publishing.

Willis, J. and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7):592–598.

Yamamura, E., Hayashi, R., Tsutsui, Y., and Ohtake, F. (2018). Racers attractive looks, popularity, and performance: How do speedboat racers react to fans expectations? Discussion Paper 18-28, Graduate School of Economics and Osaka School of International Public Policy (OSIPP) Osaka University.