Validating the win-rate of heroes in Dota 2 using instrumental variable estimation

David van der Maas
Supervisors: Jesse Krijthe, Rickard Karlsson
EEMCS, Delft University of Technology, The Netherlands

June 19, 2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

This research provides additional insights in to when instrumental variable estimation is a proper method to use when investigating or removing causal effects in randomized experiments. this is done by using instrumental variables on the game Dota 2, in which win-rates of a couple heroes are validated and reviewed to test the effectiveness and usefulness of instrumental variables. The results show mixed signs, for such a complex statistical game with many confounding factors, getting a good variable to work is hard. In the end it is concluded that Dota 2 with its randomized game-mode of single draft is still a too complex experiment to retrieve unbiased results from the instrumental variable estimation method. Therefore it is not advised to use this method in a setting where the environment is is too complex.

1 Introduction

Instrumental variable estimation or IV in short, is a statistical method to leave out all bias from confounding factors that may occur a causal experiment. The usage of this method is broad but mostly used in the field of econometrics. Despite its common appearance in causal inference studies the method's capabilities and limitations are not researched thoroughly, this is because for every experiment confounding factors and their effects may change. The goal of this research is to test the usefulness of instrumental variable estimation on validating other predictive models in the game Dota 2. Dota 2 is a multiplayer online battle arena (MOBA) game in which two teams of five players face each other in battle to see which team can destroy the opponents base. In Dota 2 many heroes can be chosen by players to be played in game. Dota 2 here is chosen because the outcome of a game is highly confounded, for example by those hero selection inputs, thus making the game complex. Such level of complexity proposes challenges for the method and might lead to biased results. However if the research shows that instrumental variable estimation can be used with such complex experiments, it might also be proven useful in other fields such as complex medical experiments or biased population models.

The exact question that will be answered is how useful instrumental variable estimation is by validating the win-rate of some heroes in Dota 2. For this research the hero Viper was initially chosen as Viper is a popular and easy to play hero which has great benefits when using IV. The general conclusion will argue whether IV is useful in these complex causal experiments.

The final result showed that instrumental variable estimation is a proper method for validation. In short a model was built using the instrumental variables method and the win-rate of Viper was compared to the publicly available one. The outcome of the publicly available win-rate was almost identical to the win-rate calculated with the IV method. This shows that the win-rate when picking Viper is not confounded much. However, another hero, Meepo, was calculated as well since this hero is really hard to play. Results showed that there was significant differences in results between the publicly available win-rate and the calculated one. Thus concluding Meepo's win-rate is heavily confounded. Finally, the IV method's results were compared to another method which also showed significant differences in results, therefore one can argue against the use of IV in such complex cases.

2 Methodology

In order to answer the research question, research into the various topics had to be conducted. First of all, instrumental variable estimation. Since this is also a subject covered in the econometrics studies, books like [3] were the main source of research into this method. Also the website on Causal Inference for the Brave and True [1] were helpful with programming and better understanding of the method.

Secondly after the research on the method was conducted, data had to be acquired from the game Dota 2 using the OpenDota API and Python scripts. This data had to be specifically chosen in order to be appropriate for the tests and research after. Firstly the games of course had to feature the hero Viper, also it was required that the game mode in which these games are played is pseudo-random. Explanation on why this is necessary will follow in section 3.

After the game data had been collected, it was possible to establish a predictive model which was not influenced by causal effects in the game but merely affected by picking Viper in a game instance. With this model another predictive model could be validated by comparing them, in this case a website on which win-rates of various heroes were displayed was used as a baseline predictive model for Vipers win-rate. Comparing the results and reasoning about why results may be different will conclude if instrumental variable estimation can be effectively used to remove unwanted bias in such complex experiments as Dota 2 or also in more general concepts.

3 Validation of predictive models using IV

Instrumental variables is used to take away bias in a causal experiment. This can be done by determining an instrumental variable which only affects the treatment of the experiment and is not confounded by any other factors. Then the treatment will influence the outcome which can be measured. A small diagram can be seen in figure 1.

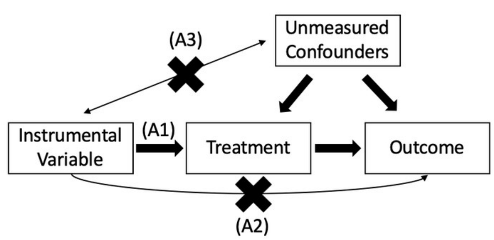

Figure 1: Diagram of instrumental variable assumptions. [4]

This figure gives a quick overview of how IV's are used. Here the chosen instrumental variable only influences the treatment through A1 (assumption 1). Also the variable does not impact the outcome directly but only through the treatment which is depicted by A2. And last the unmeasured confounders do not have impact on the variable itself and vice versa, A3.

To relate this back to Dota 2 and validating Vipers win-rate, first a suiting instrumental variable has to be chosen. Since this method works best in semi-random instances the game-mode Single Draft (SD) in Dota 2 is a proper way to introduce this method. In short SD is a game-mode where every player gets to choose their hero from a pool of 3 heroes while the others are unselectable. Also players can not see their own team's selections. Using this game-mode the reduced selection of heroes can be used as an instrument for the experiment.

Now given the instrument, one must argue in favour of this variable. Meaning are all assumptions A1, A2 and A3 true?

3.1 Assumptions

A1. Given the reduced picks in SD as variable, does this influence picking Viper, the treatment? It is possible to argue in favour of this assumption by looking at the pick-rate of Viper when a player is able to select him. In the normal game-mode Viper has a pick-rate of 10.36%. In SD the pick percentage of Viper when he is amongst the 3 heroes that are available is 49.39%. Therefore choosing this variable will indeed influence the treatment.

A2. Do the reduced picks affect the outcome of a game? To state the obvious, no. Since every player has this reduced hero pool the game is not in favour of any of the two teams. Thus the restricted hero pool does not influence the outcome in any other way than through the hero selected.

A3. Here one must reason about the instrumental variable not having any effect on any of the confounders and vice versa. Most of the confounders can be considered trivial as why they have no influence, for example skill level, unfavorable match-ups etc. However some can reason the confounding factor of team building might influence the variable. Fortunately in SD a player can not see the selection made by a teammate, therefore team building is not something players can do easily. Players can however communicate with each other in team chat which hero they chose. But due to time constraints and the intention of the game-mode to have fun playing this would be a very weak influence on the instrumental variable.

3.2 Calculation

To measure the actual win-rate when viper is picked without all bias by confounding factors one must calculate the effect of the variable on the treatment (1st stage) and then the treatment of the outcome given the variable (reduced form). To measure these effects different regression models can be used, in the instance of this experiment the Ordinary Least Squares (OLS) method was used. Another commonly used option is Two-Stage Least Squares (2SLS), however 2SLS is always biased in comparison to OLS [2] which makes it for the binary data used better and more accurate to use OLS.

$$ATE = Reducedform/1stStage \tag{1}$$

The equation 1 states that the unbiased IV estimate of the average causal effect is equal to the reduced form divided by the 1st stage. One thing to note is that if the value of the 1st stage is small, the outcome will most likely be incorrect or even pseudo-random. This is because the 1st stage describes the effect of the variable on the treatment, which in that case is very weak, thus having a weak instrument.

4 Experiment and Result

4.1 Experiment

As stated in section 3.2 the calculation of this method is relatively easy. First data from the Open Dota API has to be fetched and parsed into useful data. In this case only the games from SD are relevant in which Viper was able to be picked by someone. Then the influence of the variable on the treatment was done by calculating how many games Viper was expected to be in versus the number he was actually in. This turned out to be the 49.39%. Thus the effect was an increase in pick-rate of 39.03% since his pick-rate in normal games is 10.36%. Then the effect of the treatment on the outcome has to be calculated. Using the OLS method from the statsmodels library in Python a value for the reduced form was obtained. By dividing the two values the (dis)advantage of picking the hero Viper in the SD format can be calculated.

4.2 Results

On average over 13796 games analyzed Viper got a negative score of -1.27%. This means a player who picks Viper has a slight disadvantage of winning the game. According to websites that track performance of heroes, Viper has an average win-rate of 49.03% which means they calculated a disadvantage of -0.97%. To further test the abilities of the IV method, another hero called Meepo was also tested in the exact same way. In this case the method calculated a disadvantage for picking Meepo at -2.78%, however the performance tracking website shows he has an average win-rate of 53.71%. Which calculates to a 3.71% advantage when picking Meepo. Testing this hero aswell was a crucial part to reason about the usefulness of IV's for Dota 2, since in this case the method got a significant other result than the predictive model used. Additionally the variances for the results were also calculated to give a better understanding. Below is a table containing the results.

	Actual Win-Rate	IV's Win-Rate	Variance
Viper	49.03%	48.73%	0.24%
Meepo	53.71%	47.22%	0.49%

5 Responsible Research

For this research data from Dota 2 matches had to be analyzed. This also means user data and other personal information was acquired to do the experiment. However Dota 2 has made its data publicly available and players also consent and acknowledge that this data can be collected by anyone. It also has to be said that no truly personal information can be acquired through the use of the Open Dota API. Merely usernames, ranks and in-game preferences can be acquired. This is to be believed to be of no harm to any of the players involved.

To reproduce the results one can run the code on their own data-set. The algorithm to retrieve data is also included in the project. Furthermore, since Dota 2 is a game with updates almost weekly, checking the validation models found online must be done to percieve the information correctly.

6 Discussion

As seen from the results of the experiment the instrumental variable estimation method got an average win-rate for Viper so close to the original one on normal games indicates that the win-rate is unbiased, and Meepo to be heavily confounded. However the assumption that the confounding factors in the game do not have any impact on the variable or vice verse is a hard to reason statement. The game of Dota 2 is of such complexity and includes so many different factors that can alter the outcome of a game that reasoning for a perfect variable which will hold for all assumptions A1, A2 and A3, is almost impossible. To find a variable which suits this problem perfectly is therefore hard or even impossible. In this experiment the result was actually in favour of the chosen variable but far more testing has to be done in order to fully validate these results.

Since Meepo was an interesting hero to review and analyze, the comparison between the IV's calculated disadvantage and the disadvantage calculated by a full randomization method (S. Avgousti, 2022) was made. Using the full randomization method Meepos disadvantage was calculated at around -12% while the IV method showed a disadvantage of -2.78%. Since in theory these numbers should be somewhat the same it is worth discussing potential violations of assumptions in the IV method. In the case of choosing the bans or picks as variable it is noticed that players who have past experience with a specific hero, tend to choose it more often. This however does mean that assumption A3 as seen in figure 1 could be violated and therefore produce biased results. One additional note is that the instrument used was less significant to picking Meepo as it was to picking Viper. This could potentially invalidate the results gotten from the IV method when applied to picking Meepo.

7 Conclusions and Future Work

This research looked into how useful instrumental variable estimation can be for validating other predictive models. To do this the win-rate of the hero Viper in Dota 2 was modeled using this method and compared its result to the publicly available win-rate. The results of the IV method showed a really similar win-rate for Viper which shows that confounding factors have little to no impact when playing Viper in regular matches. However when analyzing Meepo, the difference in win-rate was significantly different. This can mean that Meepos win-rate is heavily confounded. As Meepo requires a player to have immense skill and prior expertise on the hero to be played properly, this will most likely be the reason for his confounded win-rate. Since in SD players only have 3 picks, means players with some experience with Meepo will more likely pick him than players who have little to no experience. However in Meepos case, some experience is not enough to play him properly, therefore having a slight disadvantage in the end.

Comparing the results to the fully randomized experiment as done in 6. One can strongly argue against the use of instrumental variables to be used in such complex experiments. In this case Dota 2 is far to complex to come up with a suitable variable that does not violate any assumptions. The game also has limited game modes from which data can be gathered. These game-modes can make the instrument weak or strong depending on the treatment, which leads to biased results as seen in this research. Overall, the IV method is a strong method for causal experiments, but is not suitable for high impact choices players make based on skill and prior knowledge, therefore not being totally random.

To improve the reliability of this research one can test the outcome of the experiment more thoroughly and can possibly model other heroes as well. This is needed to fully confirm that the IV method can be used for these complex settings as Dota 2.

References

[1] Matheus Facure Alves. Causal inference for the brave and true, 2022.

[2] Jason Ansel, Han Hong, and Jessie Li. Ols and 2sls in randomized and conditionally randomized experiments. *Jahrbücher für Nationalökonomie und Statistik*, 238(3-4):243–293, 2018.

[3] Roger Bowden and Turkington Darrel. *Instrumental variables.* Cambridge university press, New York, 1984.

[4] Michael Johnson, Jiongyi Cao, and Hyunseung Kang. Detecting heterogeneous treatment effect with instrumental variables, 08 2019.