Fifty Cents for Your Soul: Does Sponsorship Acquisition Change Online Creator's Content?

Felix Nguyen*

May 12, 2022

Abstract

In this research, we address the scarcely studied question of whether the transition from gratuitous hobbyists to sponsored content creators can lead to changes in the types and quality of content generated, as well as audience responses to these changes. Combining Variational Autoencoder based representation learning with causal inference methods, we examine the treatment effect of sponsorship acquisition on future content. We further investigate how the changes in each dimension of the content correspond to changes in audience reaction and sentiment. Through this research, we contribute to the nascent literature on content creator and influencers, and help brands decide when they should utilize this channel.

Keywords: Influencer, Content Creator, Reputation, Representation Learning, Causal Inference.

*PhD Student in Marketing - Emory University - pnguy38@emory.edu

1 Introduction

Online content creation has transformed in recent years from a casual, spontaneous hobby into a respectable full-time career choice among the younger generations. In a 2019 survey by Harris Poll on behalf on Lego, over 30% of children age 8-12 in the United States want to become a Youtube content creator[1]. This is also not all wishful thinking. Industry surveys have shown that online content creators wield a tremendous level of influence on what younger audience watch, what they talk about, and what they buy. According to a survey of more than 2,000 participants by *Morning Consult*, 24% Gen Z women and 16% Gen Z say influencers are where they most often learn about new products to buy, and 88% have bought a product they learned about first on social media[2]. In 2017, business magazine *Forbes* proclaimed that Youtubers are now more influential with Millennials and younger generations that traditional celebrities[3].

Catching on to this trend, in recent years firms have been allocating more and more of their marketing budget into sponsoring online content creators. In the span of just five years, the influencer marketing industry grew ten fold, from a mere $1.6 billion in 2016 to over $14 billions in 2021[4]. According to a survey in 2022 of over 5,000 marketing agencies, brands, and other industry professionals by *Influencer Marketing Hub*, around 75% brand marketers intend to have dedicated budget for influencer marketing in 2022. Additionally, total sales globally associated with social media and online content reaches approximately $958 billion. According to the same report, more than 50 million people around the world consider themselves online content creators. However, the brand-influencer relationship is not a one-way streat, as more than 77% of content

[1] https://www.businessinsider.com/american-kids-youtube-star-astronauts-survey-2019-7

[2] https://morningconsult.com/wp-content/uploads/2019/11/The-Influencer-Report-Engaging-Gen-Z-and-Millennials.pdf

[3] https://www.forbes.com/sites/under30network/2017/06/20/why-youtube-stars-influence-millennials-more-than-traditional-celebrities

[4] https://www.statista.com/statistics/1092819/global-influencer-market-size/

creation revenue still come from brand deals (as illustrated in **Figure 1**), despite the large amount of ad revenue sharing by Youtube. All of these statistics demonstrate the growing importance of influencers/online content creators as a marketing channel, and along with that, the necessity of understanding the complex relationships between sponsors, content creators, and their audience.

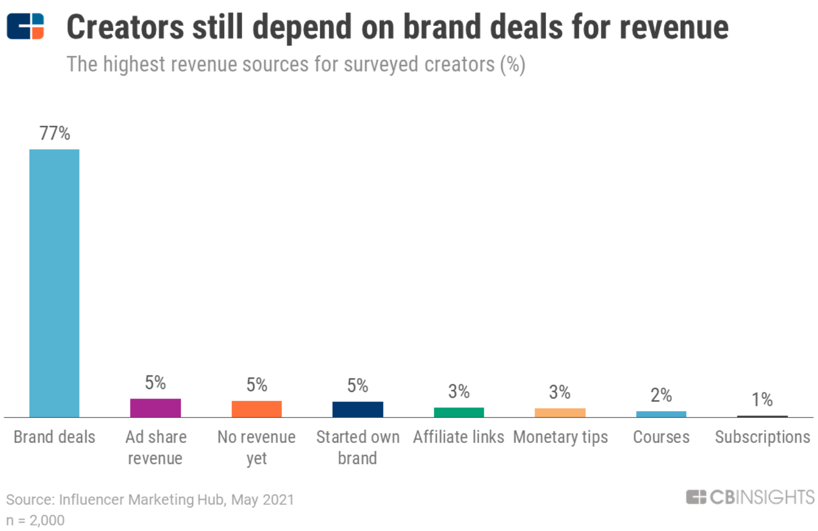

Figure 1: Content Creators Sources of Revenue

This new and exciting area of research has been receiving more and more attention in academic literature in recent years. For example, applying transfer learning to publicly available data on YouTube influencer videos, Rajaram and Manchanda (2020) find that mentioning brands early on in Youtube videos lead to a significant increase in attention to the brand but a significant decrease in sentiment expressed towards the video. In a study with data from another popular video platform, Yang et al. (2021a) show that timing and intensity of in-video sponsored content are highly predictive of its effectiveness, and propose a simple "motion" metrics constructed from Convolutional embeddings of video content to represent this relationship. With evidence from yet another popular video platform, Twitch, Huang and Morozov (2022) find that organic live streams of video games only marginally increase the number of concurrent players in these games, and the sponsor solicited streams are even less effective. The common theme between current empirical studies is that they mostly focus on the structure or timing of sponsored content, and how those characteristics can determine the effectiveness of influencer marketing.

Yet, recent and ongoing studies from a more theoretical perspective have highlighted the importance of the reverse direction of sponsor-content creator dyadic relationship, namely how brand sponsorship impacts influencers and the content they create. Pei and Mayzlin (2021) examine the tripartite between firm, influencer, and audience, and show that if the audience prior belief in the brand is low, it is optimal for firms to only loosely affliate with the influencer, and vice versa. Mitchell (2021) builds a dynamic model of relationship between an influencer and their follower, and demonstrates that this relationship follows a cyclical "reap and sow" process, in which the influencer vacillates between giving original and unbiased advice to improve trust, and spending that trust by monetizing the opportunity to advise through sponsorship.

In a similar vein, Fainmesser and Galeotti (2020) develop a model in which influencers trade off the increased revenue they obtain from paid sponsorship with the negative impact that these have on their followers' engagement and, therefore, on the price they receive from sponsors. All of these theoretical models assume that sponsored content will reduce audience trust and the influencer's reputation, and the influencer has to solve this trade-off problem. Yet, no empirical studies have examined this effect of sponsorship on influencers and the trade off mechanism, and through this research, we aim to close this gap between theory and empirical evidence.

Potential reasons for why sponsorship may affect the type of content created by influencers largely fall into two groups: economic reasons and psychological reasons. From an economic perspective, as Fainmesser and Galeotti (2020) and Mitchell (2021) point out, influencers must decide on their content in the context of a trade off between getting more sponsored revenue and reducing reputation from offering too much sponsored content. There are two possible approaches they can take to adress this trade-off. If the "reap and sow" cycle of Mitchell (2021) is prevalent, an influencer may to choose to "mass produce" lower quality and shorter sponsored videos to exploit the "sow" stage and maximize sponsorship revenue before resuming higher quality unsponsored content to rebuild the trust. Alternatively, if an influencer expects that the trust and reputation reduction effect may be cumulative and permanent, they may instead invest more effort into the content, creating higher quality content in order to "compensate" the audience for the sponsorship. It is also possible that both of these types of responses exist and dependent on the nature of content and characteristics of the creator.

From a psychological lens, it is possible that the transition from unpaid hobbyist to sponsored content creator triggers an "identity shift" (Savary and Dhar 2020) of the content creator. Prior research has shown that behavior and creativity can differ significantly within one person between when they assume an "amateur hobbyist" identity and when they assume a "professional" identity. Therefore, it is reasonable to expect that a newly sponsored content creator will try to generate more polished content, however they may also reduce the spontaneity and originality of the content, which is often associated with amateurism. A similar psychological phenomenon that may affect influencer creativity after getting sponsorship is "social schema" concept (Fiske and Linville 1980), in which an influencer may form a certain "recipe of success" for their videos and keep making similar content in order to keep the sponsorship, without deviating and exploring new content.

Applying state-of-the-art representation learning with Multimodal Disentangled Variational Autoencoder (VAE, Kingma and Welling 2013) to publicly available Youtube data, we transform the unstructured videos and their associated metadata (transcripts, thumbnail, audio) into interpretable dense vectors. With these measurable representations, we then proceed to estimate the causal effects of becoming a sponsored content creator on new content using Synthetic Difference in Differences method (Arkhangelsky et al. 2021), using the distance between new content latent vectors and an anchor point, or projection of the latent vectors into dimensions of interest, as the outcome variable. We can then see if these changes mediate the effect of sponsorship on audience reactions and sentiments.

This research contributes to the nascent literature on content creators and influencer marketing as discussed above, by providing one of the first empirical evidence of sponsorship's impact on influencers and their content. It also contributes to the growing literature on representation learning in Marketing (Dew et al. 2019, Yang et al. 2021b). Improving upon the Variational Autoencoder approach of Dew et al. (2019), the Disentangled VAE used in this research allows better interpretation of the underlying latent structure, enabling researchers to represent very high dimensional, unstructured data such as video content in a more convenient manner. The representation learning and causal inference framework proposed here is not limited to content creation context, and can be used to answer many more interesting questions where the outcome of interest is unstructured data. From a managerial perspective, this paper provides additional insights to marketers on what to expect when they sponsor an influencer, and why their influencer marketing campaign may be a success or failure. As proposed in Pei and Mayzlin (2021) theoretical model, the results of this study also help marketers calibrate their level of affliation with the influencers, and to work with influencers to keep producing high quality and engaging content after getting sponsored.

2 Context and Data

In this study, we make use of publicly available data from Youtube, the largest video content network in the world, and one of the major platforms for influencer marketing.

As of 2020, Youtube has more than 2.1 billion users globally, with 122+ million daily active users, and each day these viewers consumed over 1 billion hours of video content, an equivalent more than 5 billion views. According to a survey by *Morning Consult*, Youtube is consistently the most used platform across age groups between 13 and 38, with over 90& usage in all groups. Youtube is also by far the most favorite platform amongst Gen Z and Millennial men, and the second most favoritte platform amongst Gen Z and Millennial women after Instagram. On Youtuber, content creators are often categorized into groups based on followers count: Nano influencers (less than 1,000); Micro influencers (1,000 - 100,000); Macro influencers (100,000 - 1 million), and finally Mega influencers (over 1 million) (Sokolova and Kefi 2020). In this study, we will focus on the micro and macro-influencers groups, as nano-influencers are unlikely sponsored, and it would be difficult to identify the first time a Mega-influencer gets their sponsorship. Due to the large number of videos, we will have to pick one category to focus on. Potential categories include "Gaming", which is the top most followed amongst male viewers, and "Skincare", the top most followed amongst female viewers. Extant studies on online influencers have also mostly focused on these two segments.

Once we identify the influencers, their historical videos can be downloaded using the downloader tool at https://youtube-dl.org/. Additional meta data such as Video Description, Thumbnails, Comments on a video can be obtained with Youtube official API. Through Youtube API, we can also obtain transcripts of video where the content creator uploaded an official transcript. In order to measure audience reaction, we also scrape historical subscribers and viewership data from SocialBlade, a Youtube monitor service. Following Hwang et al. (2021), we use the content of the description box to identify sponsored video. In accordance with the celebrity endorsement policy of the Federal Trade Commission, content creators on Youtube must disclose whether they are sponsored in the description. Simple Natural Language Processing methods can employed to detect this, which often has "Thanks to [SPONSOR]...", "Use code...", or an affiliated URL in the content.

From the videos themselves, standard video processing methods can be employed to extract the main components: the still video frames, the audio, and the transcript if one cannot be obtained through Youtube API. From these data, we can also extract further features such as basic characteristics of the video (Length, Number of Scenes, Speaking Rate etc.); visual features (Saturation, Brightness, Contrast, Definition etc.); voice features (loudness, pitch, phonation...) and so on to use as our input. The raw data will also be used as input to our Encoder, following common transformation for each type (e.g. Convolutional Neural Network for frames, LSTM for audio data etc.).

3 Empirical Approach

Our main empirical approach follows a two-step process. First, we address the unstructured nature of online content with a deep learning model following state-of-the-art Representation Learning practice, projecting the content into an interpretable latent space such that each content, a combination of the video, transcription, description, thumbnail etc., can be represented by a dense vector, dimensions of which can be assigned meaningful interpretations. In the next step, we examine the causal effect of sponsorship acquisition on the content using Synthetic Difference in Differences method (Arkhangelsky et al. 2021) with the content latent vectors as the main outcome of interest. We can then examine if these changes mediate or moderate audience reception to the transition from hobbyists to sponsored content creators.

3.1 Learning Content Representation with GDM-VAE

3.2 Classical VAE

The first challenge we have to tackle in the empirical analysis is how to combine a wide array of unstructured data extracted in the previous section, ranging from images of videoframe, audio data, text of the transcript, to the thumbnails or the video description. Traditionally, in Marketing and elsewhere, the main approach is to manually extract a limited number of properties that one want to examine. With recent advances in machine learning and especially Deap Learning, however, we no longer have to go through such painstaking efforts. Instead, "Representation Learning" methods can be used to seamlessly extract the representation of the data in an unsupervised or semi-supervised manner. In this research, our representation learning method of choice is through a Variational Autoencoder.

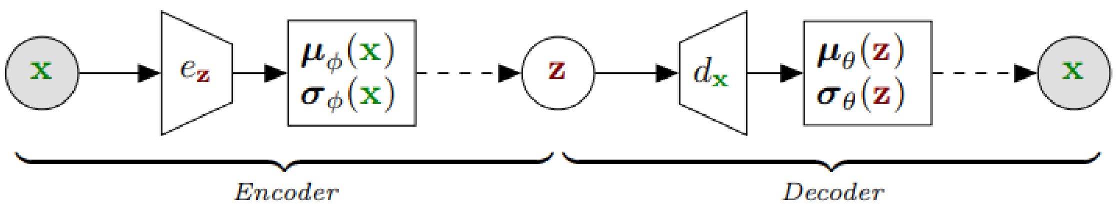

Figure 2: Basic Schematic of a Variational Autoencoder

Variational Autoencoder (VAE, Kingma and Welling 2013) is a form of Generative Deep Learning models, which try to learn the data generating process and can be used to generate new data, as opposed to Discriminative Learning model which can only be used to classify or predict without learning the actual DGP. A basic VAE consists of three main integrated components: a latent structure that is assumed to be the main factors that govern the data generating process, a generative model called the *Decoder* for the observed data that is conditional on the latent variables, and an *amortized* variational distribution that approximates the posterior distribution of the observation-specific latent variables called the "Encoder". The Encoder and Decoder are jointly estimated during training phase. This overall structure is illustrated in **Figure 2**.

Specifically, the generative "Decoder" approximate the probability distribution of the observed input tensor $\boldsymbol{x}_i$ for observation i, as a function of the latent tensor $\boldsymbol{z}_i$. This approximation is done with a multilayered neural network, governed by the parameters

$\boldsymbol{\theta}$. Formally, the decoder is defined by:

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_i, \boldsymbol{z}_i) = p_{\boldsymbol{\theta}_x}(\boldsymbol{x}_i|\boldsymbol{z}_i)p_{\boldsymbol{\theta}_z}(\boldsymbol{z}_i)$$

In which the prior of the latent space is assumed to be Gaussian:

$$p_{\boldsymbol{\theta}_z}(\boldsymbol{z}_i) \sim \mathcal{N}(\boldsymbol{z}, \mathbf{0}, \boldsymbol{I})$$

In a traditional VAE, the parameters of the latent space $\boldsymbol{\theta}_z$ is empty and the prior is simply $p(\boldsymbol{z}_i) \sim \mathcal{N}(0, \boldsymbol{I})$. Next, the conditional distributions $p_{\boldsymbol{\theta}_x}(\boldsymbol{x}_i|\boldsymbol{z}_i)$ is modeled with neural network called the *decoder network*. The output of this network is the parameters of $p_{\boldsymbol{\theta}_x}(\boldsymbol{x}_i|\boldsymbol{z}_i)$. Distributional assumptions vary with the type of x_i, however generally in the exponential family. For example, with a continuous x_i (e.g., values of each pixel in an image tensor, waveforms of audio, brightness of the video, values of word embeddings etc.), Gaussian distributions are typically assumed. On the other hand, categorical x_i, such as labeled attributes, influencer emotions, etc. are often modeled with a categorical or multinomial distribution.

The "Encoder" part approximate the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{z}_i|\boldsymbol{x}_i)$ via *amortized variational inference*, estimating aa *approximate posterior* distribution $q_{\boldsymbol{\phi}}(\boldsymbol{z}_i|\boldsymbol{x}_i) \approx p_{\boldsymbol{\theta}}(\boldsymbol{z}_i|\boldsymbol{x}_i)$, the vector of weights and biases $\boldsymbol{\phi}$ is amortized, i.e. shared across observations, for scalable inference. In conjugation with the prior, the amortized posterior is often assumed to be Gaussian, with the mean and covariance parameters estimated through multilayered neural network.

3.3 Guided Disentangled Multi-modal VAE (GDM-VAE)

One weakness of the classical Variational Autoencoder model is that the latent space has been shown to be highly "entangled", i.e. the dimensions are highly correlated (Chen et al. 2018, Alemi et al. 2018), which stemmed from the fact that there is a family of models with identical ELBO (the objective of VAE and other variational deep learning models), but different quantitative and qualitative characteristics, and thus the latent space we learn is often a mix of these (so called "broken ELBO problem"). This makes the problem of identifying and interpreting the meanings of the latent space an all but impossible task, and most of the time the dimensions are largely meaningless on their own. This is different from low dimensional, classical latent factor models such as Principal Components Analysis or Latent Dirichlet Allocation, where we can often assign some labels to the extracted topics or dimensions. In order to address this gap, recently there has been a new stream of literature on "disentangled" VAE (Higgins et al. 2016, Kim and Mnih 2018, Mathieu et al. 2019), which modify the objective function to make sure the latent space consists of orthogonal dimensions that may be interpretable. Another recent stream of work by Khemakhem et al. (2020) and Sorrenson et al. (2020) takes inspration from nonlinear ICA models and suggests making latent space interpretable by conditioning the latent prior $p_{\boldsymbol{\theta}_z}(\boldsymbol{z}_i)$ on the information of sets of known labels $\boldsymbol{u}$ to get $p_{\boldsymbol{\lambda}}(\boldsymbol{z}_i|\boldsymbol{u})$ to "guide" the latent space.

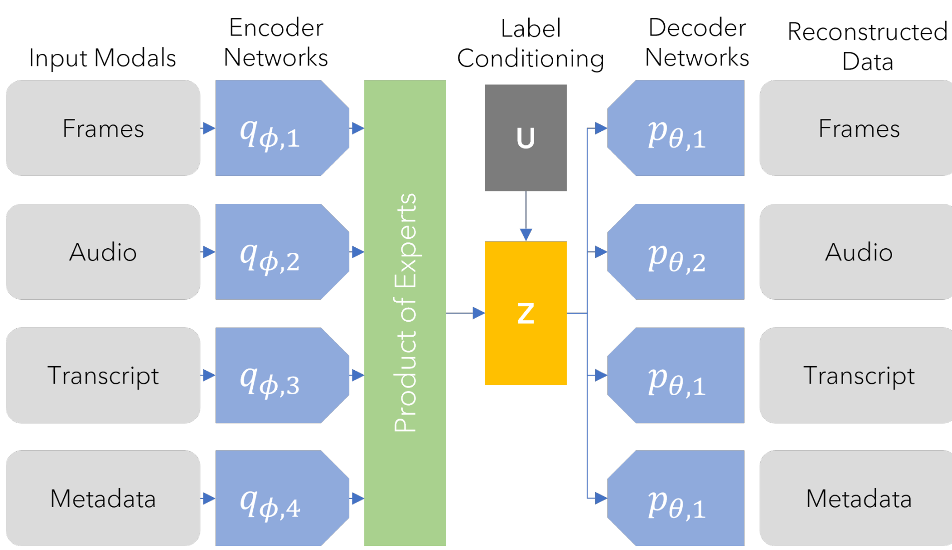

Figure 3: GDM-VAE Model Schema

In this paper, we propose a combination of two approaches above, with a Guided

Disentangled Multi-modal VAE (GDM-VAE) model. In this model, we condition the prior of our latent space on a set of labels of each observation, with the conditional prior following a distribution in the exponential family:

$$p_{\boldsymbol{T},\boldsymbol{\lambda}}(\boldsymbol{z}|\boldsymbol{u}) = \prod_{i=1}^{m} p(\boldsymbol{z}_i|\boldsymbol{u}) = \prod_{i=1}^{m} \frac{Q_i(\boldsymbol{z}_i)}{Z_i(\boldsymbol{u})} \exp\left[\sum_{j=1}^{k} T_{i,j}(\boldsymbol{z}_i)\lambda_{i,j}(\boldsymbol{u})\right]$$

In which Q_i is the base measure, $\boldsymbol{T}_i = (T_{i,1}, ..., T_{i,k})$ are the sufficient statistics, $Z_i(\boldsymbol{u})$ is the normalizing factor, $\boldsymbol{\lambda}_i = (\lambda_{i,1}(\boldsymbol{u}), ..., \lambda_{i,k}(\boldsymbol{u})$ are the natural parameters, and k is the pre-defined number of sufficient statistics. In our main analysis, we use Gaussian distribution following convention, but this can be any distribution from the exponential family. Similar to the Encoder and Decoder networks, this conditional distribution is estimated with a multilayered neural network that gives the mean and the variance parameters as output. The label information could come from self generated "tags" of the videos by the content creators, the category of games/products being discussed, or additional features extracted from the data as discussed above such as sentiments, facial expression, topics of the video.

Next, adapting the objective function from disentangled model β-TCVAE (Chen et al. 2018), we can write down the $\beta-$ disentangled ELBO with labels information, of N modalities and K dimension latent structure as follow:

$$\begin{aligned}\mathcal{L}_{GDM-VAE} = E_{q_\phi}\left[\sum_{i=1}^{N} \lambda_i \log p_\theta(x_i|\boldsymbol{z},\boldsymbol{u})\right] - KL\left[q_\phi(\boldsymbol{x},\boldsymbol{z})||q_\phi(\boldsymbol{z})p(\boldsymbol{x})\right] \\ - \beta KL\left[q_\phi(\boldsymbol{z})||\prod_{k=1}^{K} q_\phi(z_k)\right] - \sum_{k=1}^{K}\left[q_\phi(z_k|\boldsymbol{u})||p_{\boldsymbol{T},\boldsymbol{\lambda}}(z_k|\boldsymbol{u})\right]\end{aligned}$$

Here, λ_i helps balancing reconstruction across different modalities, β is a penalty parameter to "disentangle" the dimensions by penalizing the "total correlation" between latent dimensions $KL\left[q_\phi(\boldsymbol{z})||\prod_{k=1}^{K} q_\phi(z_k)\right]$. This would be the main hyperparameter that requires tuning in our training process. Our objective serves a similar purpose as the semi-supervised objective of Cheng et al. (2022)'s InnoVAE model, however here we are not limiting some dimensions to be exactly the same as the labels, instead passing on the labels information to automatically separate the dimensions. Thus, for example, if we include the sentiments extracted from text transcript, the model can discover related sentiment dimensions from video/audio, or sentiments can be split into multiple dimensions (sad vs. happy, extreme vs. mild). Additionally, by penalizing only the Total Correlation KL instead of the full KL, we ensure disentanglement without sacrificing reconstruction quality.

3.4 Causal Inference with Synthetic DiD

3.4.1 Synthetic Difference in Differences

Once we have obtained the representation z_i of the content, it is straightforward to apply common causal inference techniques with this representation as the outcome, and the sponsorship acquisition event as the treatment. In this research, we use Synthetic Difference in Differences approach by Arkhangelsky et al. (2021), a combination of popular case study method Synthetic Control with common panel data causal inference method Difference in Differences. This approach relaxes the parallel trends assumption in classical Difference in Differences by constructing a "synthetic control" from a convex combination of actual control units' outcome of interest that are constrained to the parallel trend assumption. Additionally, instead trying to construct a counterfactual version of the treated unit as the synthetic control, in this method we only need to construct the counterfactual "control" that then can be used in normal Difference in Differences setup. This enables us to apply the method to contexts where there are more than one treated unit, which is not possible with classical Synthetic Control method.

Mathematically speaking, in Synthetic DiD, we solve this optimization problem:

$$\underset{\tau,\mu,\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\min}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\mu-\alpha_i-\beta_t-\tau W_{it})\omega_i\lambda_t\right\} \tag{1}$$

In which, μ, α_i, β_t are the usual two-way fixed effects in a panel DiD setup, τ is the average treatment effect, $\boldsymbol{W}$ is a block matrix of treatment status (i.e. being sponsored), with $W_{it} = 1$ if i is in the treatment group and t is post-treatment. We can see that these elements are similar to a standard two-way fixed effects regression. The main difference here is unit weights ω_i and time weights λ_t, which are optimized to ensure that the synthetic control satisfies parallel trends (See Arkhangelsky et al. (2021) for details on calculating these weights). Hence, this can also be interpreted as a time-and-unit weighted two-way fixed effects estimator. Figure 1 from Arkhangelsky et al. (2021) presented here (as **Figure 4**) illustrates the differences and similarities between Synthetic DiD, DiD, and Synthetic Control.

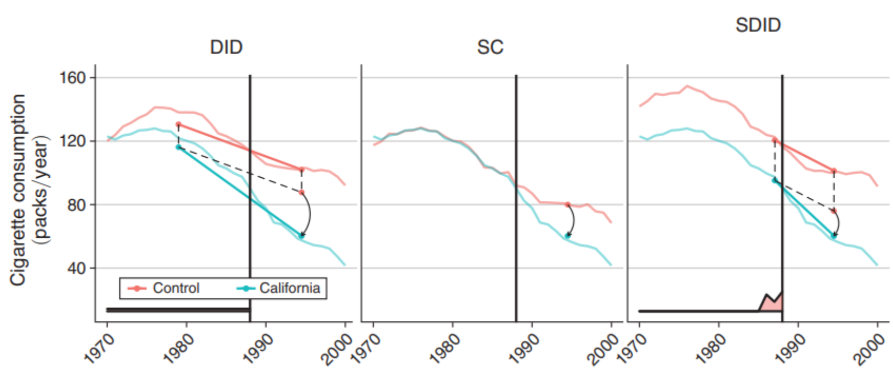

Figure 4: Synthetic DiD, DiD, and SC applied to California Smoking Cessation

3.4.2 Outcome Variables

Now for the outcomes Y_{it}, we can first construct a general measure of how much the content of a creator changed after they receive sponsorship. Let:

$$\boldsymbol{z}_{it} = \frac{1}{N_{it}} \sum_{j=1}^{N_{it}} \boldsymbol{z}_{ijt}$$

be the average vector of the latent presentations of all content $j \in N_{it}$ produced by creator i during time period t, which we can extract from the trained Encoder networks of GDM-VAE. We can then operationalize "change" by calculating the distance between $\boldsymbol{z}_{it}$ and:

$$\boldsymbol{z}_0 = \frac{1}{N_0} \sum_{j=1}^{N_0} \boldsymbol{z}_j$$

Which is a "baseline" vector that can be either the average of latent presentation of all content before the creator getting a sponsor, or of recent periods before getting a sponsor. We can think of this vector as the presentation of the typical type and style of content produced by the creator when they were a hobbyist. The distance can be a simple L2 distance measure:

$$D_{it} = ||\boldsymbol{z}_{it} - \boldsymbol{z}_0||_2$$

Alternatively, we can use the Cosine similarity:

$$D_{it}^{cos} = \frac{\boldsymbol{z}_{it} \cdot \boldsymbol{z}_0}{||\boldsymbol{z}_{it}||||\boldsymbol{z}_0||}$$

Thus, once these measures are used as $\boldsymbol{Y}$ in the Synthetic DiD framework, the treatment effect we capture represents the change in the style and substance of content produced by the creators after receiving sponsorship in comparison to what it would have been without sponsorship, and after accounting for creators inherent heterogeneity and time varying shifts and trends in content (e.g. changing viewers' taste, platform policy changes).

Another property we can look at with the latent presentations is the originality of the content. Let:

$$\mathcal{S}_{t-1} = \text{span}\{\boldsymbol{z}_{1,t-1}, \boldsymbol{z}_{2,t-1}...\}$$

That is, $\mathcal{S}_{t-1}$ is the smallest linear subspace that contains the set of latent vectors of content from all creators in period t_1. Furthermore, let $Pr_{\mathcal{S}}(.)$ be the projection operator of a vector to this subspace. Then, we propose this simple measure of originality:

$$Originality_{it} = ||\boldsymbol{z}_{it} - Pr_{\mathcal{S}}(\boldsymbol{z}_{it})||$$

This captures how different the content vector is to the "content space" in the previous period, and thus can give us an idea of the level of originality of the content. To put it simply, this measure is a literal parameterization of the term "out of the box" in terms of multidimensional latent space, with the box here being the span subspace $\mathcal{S}_{t-1}$.

Finally, as we can extract interpretable dimensions with the GDM-VAE model, we can simple use the coordinates of $\boldsymbol{z}_{it}$ in these dimensions as other outcomes. It is also trivial combine multiple dimensions as some convex combinations of the coordinates. Some examples could be measuring the effect of sponsorship on content quality, on product value, on sentiment etc.

Additionally, in order to tie these measures of content properties to audience reaction, we can employ two-way fixed effect regressions:

$$Y_{jit} = \beta_1 D_{jit} + \beta_2 Originality_{jit} + \sum_{k=1}^{K} \gamma_k z_{jit,k} + \eta_t + \eta_i + \varepsilon_{jit}$$

In which, Y_{jit} is a measure of audience reaction to a content j by creator i in time t such as number of views, likes, number of followers gained etc., $z_{jit,k}$ is the value of dimension $k \in \{1, 2, ..K\}$ of representative vector $\boldsymbol{z}_{jit}$, and η_t, η_i are time and creator fixed effects. We can further restrict the data to not treated periods only to avoid bias from sponsorship treatment. These regressions will give us an idea of how each content property influence audience reactions, and by establishing the relationships between sponsorship and content properties, and content properties and audience reactions, we gain deeper insights into the mechanism of why audience reactions may become unfavorable after a creator gets sponsored, as found in other recent empirical studies.

4 Conclusion

In this paper we study the effect of sponsorship on the content created by online influencers. We make novel contributions to the literature in two main ways. First, we are among the first studies to provide empirical evidence on the effect of sponsors on influencers, and the first one to focus on the content they created after sponsorship, while extant literature has mostly focused on the reverse direction of the effect of content on sponsor's marketing metrics such as sales or click-through rate. This validates the complicated two-way relationship between firms and influencers, as well as the trade-off influencers had to make when taking on sponsorship, as demonstrated in recent game-theory based analytical models.

Methodologically, we provide a general framework of performing causal inference with unstructured data such as online content as the outcomes. Our two step procedure of representation learning and causal inference is not limited to the context of this paper, but can be used to address many other interesting questions, such as peer effect on content generation, or effects of firm's policies on multi-modal user generated content (e.g. reviews with images). Additionally, the GDM-VAE model here is an improvement over prior VAE-type models in marketing literature, as it can generate more interpretable latent dimensions, giving researchers more opportunities to perform inference on the latent structure.

References

Alemi, A., Poole, B., Fischer, I., Dillon, J., Saurous, R. A., and Murphy, K. (2018). Fixing a broken elbo. In International Conference on Machine Learning, pages 159–168. PMLR.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. American Economic Review, 111(12):4088–4118.

Chen, R. T., Li, X., Grosse, R. B., and Duvenaud, D. K. (2018). Isolating sources of disentanglement in variational autoencoders. Advances in neural information processing systems, 31.

Cheng, Z., Lee, D., and Tambe, P. (2022). Innovae: Generative ai for understanding patents and innovation. Available at SSRN.

Dew, R., Ansari, A., and Toubia, O. (2019). Letting logos speak: Leveraging multiview representation learning for data-driven logo design. Available at SSRN, 3406857.

Fainmesser, I. P. and Galeotti, A. (2020). Pricing network effects: Competition. American Economic Journal: Microeconomics, 12(3):1–32.

Fiske, S. T. and Linville, P. W. (1980). What does the schema concept buy us? Personality and social psychology bulletin, 6(4):543–557.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2016). beta-vae: Learning basic visual concepts with a constrained variational framework.

Huang, Y. and Morozov, I. (2022). Video advertising by twitch influencers. Available at SSRN 4065064.

Hwang, S., Liu, X., and Srinivasan, K. (2021). Voice analytics of online influencers—soft selling in branded videos. Available at SSRN 3773825.

Khemakhem, I., Kingma, D., Monti, R., and Hyvarinen, A. (2020). Variational autoencoders and nonlinear ica: A unifying framework. In International Conference on Artificial Intelligence and Statistics, pages 2207–2217. PMLR.

Kim, H. and Mnih, A. (2018). Disentangling by factorising. In International Conference on Machine Learning, pages 2649–2658. PMLR.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Mathieu, E., Rainforth, T., Siddharth, N., and Teh, Y. W. (2019). Disentangling disentanglement in variational autoencoders. In International Conference on Machine Learning, pages 4402–4412. PMLR.

Mitchell, M. (2021). Free ad (vice): internet influencers and disclosure regulation. The RAND Journal of Economics, 52(1):3–21.

Pei, A. and Mayzlin, D. (2021). Influencing the influencers. Available at SSRN 3376904.

Rajaram, P. and Manchanda, P. (2020). Video influencers: Unboxing the mystique. Available at SSRN 3752107.

Savary, J. and Dhar, R. (2020). The uncertain self: How self-concept structure affects subscription choice. Journal of Consumer Research, 46(5):887–903.

Sokolova, K. and Kefi, H. (2020). Instagram and youtube bloggers promote it, why should i buy? how credibility and parasocial interaction influence purchase intentions. Journal of retailing and consumer services, 53:101742.

Sorrenson, P., Rother, C., and Köthe, U. (2020). Disentanglement by nonlinear ica with general incompressible-flow networks (gin). arXiv preprint arXiv:2001.04872.

Yang, J., Zhang, J., and Zhang, Y. (2021a). First law of motion: Influencer video advertising on tiktok. Available at SSRN 3815124.

Yang, Y., Zhang, K., and Kannan, P. (2021b). Identifying market structure: A deep network representation learning of social engagement. Journal of Marketing, page 00222429211033585.