How Mega is the Mega? Measuring the Spillover Effects of WeChat by Machine Learning and Econometrics

Jinyang Zheng

Michael G. Foster School of Business, University of Washington, zhengjy@uw.edu

Zhengling Qi

Department of Statistics, University of North Carolina, Chapel Hill, qizl1027@live.unc.edu

Yifan Dou

School of Management, Fudan University, yfdou@fudan.edu.cn

Yong Tan

Michael G. Foster School of Business, University of Washington, ytan@uw.edu

Abstract

WeChat, an instant messaging app, is considered a mega app due to its dominance in terms of usage among Chinese smartphone users. Nevertheless, little is known about its externality in regard to the broader app market. Our work estimates the spillover effects of WeChat on the other top-50 most frequently used apps in China through data on users' weekly app usage. Given the challenge of determining causal inference from observational data, we apply a graphical model and econometrics to estimate the spillover effects through two steps: (1) we determine the causal structure by estimating a partially ancestral diagram, using a Fast Causal Inference (FCI) algorithm; (2) given the causal structure, we find a valid adjustment set and estimate the causal effects by an econometric model with the adjustment set as controlling non-causal effects. Our findings show that the spillover effects of WeChat are limited; in fact, only two other apps, Tencent News and Taobao, receive positive spillover effects from WeChat. In addition, we show that, if researchers fail to account for the causal structure that we determined from the graphical model, it is easy to fall into the trap of confounding bias and selection bias when estimating causal effects.

Keywords: *causal inference, graphical model, app analytics, WeChat, spillover effects, machine learning, econometrics*

1 Introduction

WeChat, seemingly a messaging app, is actually more of a portal, a platform, or even a mobile operating system, depending on one's perspective (Chen, 2015). Launched in 2011, WeChat has one billion registered users and 550 million active users who open the app more than 10 times a day. Usage of the app has contributed $1.76 billion to lifestyle spending and $15.3 billion mobile data consumption in 2014, indicating its mega status in terms of smartphone usage among its users (Cormack, 2015). Industrial anecdotes related to its large scale and user engagement suggest the spillover effects of WeChat. Specifically, its intensive usage might reshape individuals' mobile usage of other apps such that apps with a higher degree of connectivity or functional complementarity to WeChat could achieve high levels of popularity and usage. This effect, however, has not been examined or measured accurately, warranting investigation of the externality of this mega-app.

Recent advancements in app analytics help researchers to understand the usage externality of apps. Ghose and Han (2014) estimate the demand of apps, given their measurable characteristics, and find measurable evidence of the use of in-app purchase design and the removal of in-app advertisements as a means to compete for market share. Other research understands the externality of app demand through special designs for the app marketplace through a rank system, as ranking naturally embeds externality. Carare (2012) Carare (2012) Carare (2012) Carare (2012) Carare (2012), who quantitatively measured users' willingness to pay for top-ranked apps, find that it is an additional $4.50 as compared to that of the same unranked app. Garg and Telang (2013) find the "bigger getting bigger" effect, specifically, that the top ranking for paid apps results in 150 times more downloads than the rest of the apps ranked in the top 200 list.

Such research, however, typically focuses on app installation as the measure of usage. Because the post-installation behaviors of users for different apps vary significantly, conditional on the installation of those apps, research is needed to further understand the externality of app usage patterns. Although there is another category of literature in the computer science field that concerns the prediction of post-installation usage patterns (Falaki et al. 2010, Tongaonkar et al. 2013, Xu et al. 2013), such research uncovers only the association rules of app usage patterns and does not provide an interpretation and measure of causality. Thus, such research is insufficient to account for the externality of an app in terms of an economic interpretation.

We address this research gap by estimating the spillover effects of WeChat usage through the use of observational data. This research objective is methodologically challenging for the following reasons. First, given the enormous size of the app market, it is difficult to identify all the apps affected by WeChat. Second, potential endogeneity issues might exist due to the uncertainty of the causal structure. Researchers who fail to account for confounders and the direction of causality might incorrectly take associations as causal effects. Both challenges are extremely difficult to address in the framework of traditional econometrics, when only observational data are available, due to the lack of a causal structure and incomplete information, such as hidden variables.

We propose to integrate a machine learning method with econometrics to identify the spillover effects of WeChat. Specifically, we introduce a Directed Acyclic Graph (DAG) and its unique representation, Completed Partially Directed Acyclic Graph (CPDAG), to characterize the underlying directed causal effect between random variables. Due to the potentially hidden variables that exist behind the observed data, we use a maximal ancestral graph (MAG) and its unique presentation, partial ancestral graph (PAG), to capture causal effects represented by observed variables. We then apply Fast Causal Inference (FCI) and Really Fast Causal Inference (RFCI) algorithms to estimate a PAG uniquely from observational data. Given the estimated PAG, we first identify the adjustment set by two kinds of recently proposed criteria: generalized back-door criterion (GBC) and generalized adjustment criterion (GAC). With the adjustment set and the condition of multivariate normal distribution, we show that the mean causal effects can be estimated quantitatively with a simple econometric linear model.

Our results show that, surprisingly, WeChat has very limited spillover effects on other apps. Only two apps, Taobao and Tencent News, receive positive spillover effects among the Top -50 apps. Our results reveal the true pattern of causality behind the association commonly observed for most of the apps, suggesting that app developers should be reserved about the connection to WeChat, as the spillover effects for most of the other apps might not be as significant as the associations with other apps. In addition, our results emphasize the advantages of using a PAG to estimate causal effects, e.g., uncovering latent confounders (identifying L in $X \leftarrow L \rightarrow Y$ by observing $X \leftrightarrow Y$), avoiding reversed causality (differentiating $X \rightarrow Y$ from $X \leftarrow Y$), and avoiding selection bias (identifying collider in $X \rightarrow Y \leftarrow Z$). We demonstrate these advantages by showing the discrepancy between causal effects encoded in the graph and those estimated with an incorrect interpretation of the causal structure or when the causal structure is unknown.

In our newly introduced method, we use several ways to rigorously evaluate the model performance. First, we test the robustness to additional information by estimating our model, using top-100 frequently used apps and top-300 frequently used apps. Second, we test our model on different weeks, including holiday and non-holiday weeks, and use different samples to ensure its stationarity longitudinally and cross-sectionally in both graphical and quantitative manners. Third, because a PAG needs to perform a conditional independence test, we check the consistency under different specifications of type-1 error levels. The results suggest a high degree of robustness.

To the best of our knowledge, this is the first application paper that integrates the most recent Bayesian network methods as FCI-PAG/RFCI-PAG (PAG estimated by FCI and PAG generated by RFCI correspondingly) and GBC/GAC with econometrics to conduct causal inference. Our research shows the strength of these methods in identifying causal relationships from observational data and suggests the feasibility of determining causal inference when an experimental setting is unavailable or costly. Note that the identification of the causal direction lies in the additional information. This approach also shows its potential in the era of big data, given the ubiquitous availability of additional information. We believe in the potential of the approach to contribute to business analytics area.

We structure our paper as follows. In Section 2, we introduce the method; specifically, we explain how to use a graphical model to represent the causal relationship of data. Given the mapping between the data and graph, we then introduce how to recover/learn causal structure from observational data graphically. We then present how to transform the information from the graph into a simple regression that can quantitatively estimate the spillover effects. We include a discussion of the relevant literature and our methods to aid readers' understanding. In Section 3, we describe the data that we use in the empirical application, and, in Section 4, we present the estimation results. We provide the robustness check in Section 5, and, in Section 6, we discuss the limitations and provide directions for further research.

2 Causal Inference by Graphical Model

A graphical model is an extremely powerful probabilistic tool for modeling the uncertainty within objects, e.g., the conditional dependence structure among random variables. Such a model can provide a clear and effective way to represent a large-scale complex system under mild assumptions. It also can provide a probabilistic inference method within an acceptable time. In addition, the presentation of a graphical model provides an intuitive understanding of the relationship among instances within a system. There are two common types of graphical models: One is Bayesian networks, which are based on directed graph, and the other one is Markov networks, or a Markov random field, which is based on undirected graph. To discover the causal relationships among instances, researchers apply Bayesian networks.

2.1 Graphical Model to Represent Causal Structure

Bayesian networks were first introduced by Pearl (1982) in the area of artificial intelligence. Later, Pearl developed a probabilistic factorization to represent the causal effect among random variables. Currently, Bayesian networks are a key area of research in machine learning and statistics. For example, as one of the most popular classification methods, Naive Bayes uses ideas of Bayesian networks.

We first introduce the basic definition of a graph. A *graph* can be represented as a pair $G=(V,E)$, where V is a finite non-empty set of vertices, and E is a set of edges formed by linking two different vertices in *V*, where there is, at most, only one edge between each pair of vertices. In general, there are four types of edges: $\rightarrow$ (directed), $\leftrightarrow$ (bi-directed), $\circ-\circ$ (undirected) and $\circ\rightarrow$ (partially directed). A *partial mixed graph* can contain all four types of edges, while a *directed graph* contains only directed ones, and a *mixed graph* can contain both directed and bi-directed edges. We have a *skeleton* of the graph by ignoring the mark of each edge. If there is an edge between two vertices, then they are *adjacent*. A *path* is a sequence of adjacent vertices. We say that a path is a *directed path* if, for every two adjacent vertices, X_i, X_j, $X_i \rightarrow X_j$ occurs. A *directed cycle* is a directed path from a vertex to itself. A directed graph *G* is called a *DAG* if it does not contain a directed cycle. Given two vertices, *X* and *Y*, if $X \rightarrow Y$, then *X* is a *parent* of *Y*. If there is a path from *X* to *Y*, then *X* is an *ancestor* of *Y*, and *Y* is *descendant* of *X*. Otherwise, *Y* is a *non-descendant* of *X*. A path $< X_i, X_j, X_k >$ is an *unshielded triple* if X_i and X_k are not adjacent. A non-endpoint vertex X_i on a path is a *collider* if the path contains $* \rightarrow X_i \leftarrow *$, where the symbol $*$ represents an arbitrary edge mark. If it is not a collider, then we call it *non-collider* on the path. A *collider path* is a path on which every non-endpoint vertex is a collider.

A *causal Bayesian network* consists of the joint probability distribution of random variables and a directed graph that encodes the causal relationship. Each vertex in *V* represents a random variable. Let *P* be the joint probability distribution of the random variables in *V*, and *G* = (*V*, *E*) is a DAG; we then define (*G*, *P*) as a *Bayesian network*. A Bayesian network is a causal Bayesian network if the graph is interpreted causally. The graph and probability are connected through the following two fundamental assumptions (Neapolitan et al., 2004; Pearl, 2011; Scheines, 1997): Markov condition and faithfulness condition.

Markov condition: A DAG and probability *P* satisfies the Markov condition if and only if, for every random variable *X* in *V*, *X* is independent of $V \setminus \{parents(X) \cup Decendant(X)\}$. If the graph satisfies the Markov condition, it means that, for each variable $X \in V$, *X* is conditionally independent of the set of all its non-descendent *ND*(*X*), given that the set of all its parents *Parents*(X), that is:

$$P(X, ND(X) \mid Parents(X)) = P(X \mid Parents(X)) P(ND(X) \mid Parents(X)) \qquad (1)$$

This condition not only interprets a DAG as a causal hypothesis but also provides tools for the practice of constructing a Bayesian network by diagnosing such statistical hypothesis testing, which we will discuss later.

Faithfulness condition: If all the conditional independence relations in *P* are entailed by the Markov condition applied to *G*, then it is faithful. When these two assumptions are satisfied, a DAG characterizes conditional independence relationships in *P* via *d-separation* (Spirtes et al., 2000).

A DAG is not fully identifiable. Several DAGs may encode the same conditional independence relation. Those DAGs form a Markov equivalence class that can be uniquely represented by a CPDAG. A CPDAG contains the same skeleton and collider structure as DAG(s). Any edge $X_i \to X_j$ in a CPDAG means $X_i \to X_j$ in every DAG in the Markov equivalence class, while an edge $X_i \circ\!-\!\circ X_j$ represents uncertainty in the Markov equivalence class, suggesting that both $X_i \to X_j$ and $X_i \leftarrow X_j$ occur in some DAG(s).

A DAG can represent a causal structure fully in the condition that we have all vertices observed. This condition, however, is barely satisfied when we try to recover the causal structure from data due to the existence of hidden variables or selection variables. Failing to satisfy the condition may cause estimation bias and incorrectly signal a causal relationship. To allow latent variables and selection variables, one can transform the underlying DAG with hidden variables and selection variables into a unique *maximal ancestral graph* (MAG) based only on the observed variables (Richardson and Spirtes, 2002). Recall that a mixed graph has four types of edges. Here, *ancestral graph* is defined as a mixed graph *G* without directed cycles and without almost directed cycles, where *almost directed cycles* occur if $X \leftrightarrow Y$ and $Y \in Ancestor(X)$.

A MAG is characterized by every two non-adjacent vertices *X* and *Y* as conditionally independent, given a subset of the remaining observed random variables. In particular, a MAG that contains a tail mark $X -\!* Y$ means that *X* is an ancestor of *Y in all DAGs represented by this MAG*. If $X *\!\to Y$ in *M*, then, in every DAG represented by *M*, *Y* is not an ancestor of *X*. In addition, the MAG of a causal DAG is called a causal MAG. The conditional independence relationship in a MAG is encoded by m-separation, which is a generalization of d-separation in a DAG (Zhang, 2008). Every pair of two non-adjacent vertices in *M* are m-separated by a subset of the remaining vertices.

With respect to identification, similar to a DAG, several MAGs may encode the same conditional independence structure and form a Markov equivalent class. Those MAGs could be uniquely represented by a PAG. Like a CPDAG, a PAG has the same skeleton as every MAG in the Markov equivalent class. The relationship between MAGs and a PAG is similar to that between DAGs and a CPDAG. If $X_i -\!* X_j$ stays constant in every MAG of Markov equivalent class, it will also present as $X_i -\!* X_j$ in a PAG. If there is an uncertain circle mark in a PAG, such as $X_i \circ\!-\!* X_j$, then the Markov equivalent class of MAGs will contain at least one $X_i -\!* X_j$ and at least one $X_i \leftarrow\!* X_j$.

2.2 Recovering Causal Structure

In Section 2.1, we showed that the causal structure can be represented by a graphical model. Using a graphical model to conduct causal inference thus consists of two stages. In the first stage, we learn about the causal structure graphically from observational data by recovering a CPDAG (in a hidden-variable-and-selection-variable-free context) or a PAG, which represents all identifiable causal relationships. The second stage involves parameter learning, in which we estimate the causal effects quantitatively based on the graphical structure of Stage 1. We discuss these two steps in detail in the following sections.

2.2.1 Stage 1: Recovering Causal Diagram / Learn the Graph

In the literature, there are two approaches to this stage. The first approach is the search-and-score approach that is based on a search procedure and the scoring metric. In this regard, it is to search the best networks by optimizing a predefined scoring metric. Well-known scoring functions include K2-CH metric (Cooper and Herskovits, 1992), chain-based scoring (Kabli et al., 2007), BDeu (Buntine, 1991), Minimum Description Length (Heckerman et al., 1995), and BIC (Schwarz et al., 1978). Because a direct search across all possible graphs is computationally infeasible due to the fact that the number of graphs grows exponentially with the number of random variables, efficient searching or optimizing methods, such as the K2 algorithm (Cooper and Herskovits, 1992), Hill Climbing (Tsamardinos et al., 2006]), Genetic Algorithm (Larrañaga et al., 1996), Simulated Annealing (Wang et al., 2004), Particle Swarm Optimization (Cowie et al., 2007), and Ant Colony Optimization (De Campos and Huete, 2000; Campos et al., 2002), have been proposed to approximate the optimal solutions.

The second approach is the constraint-based learning method that discovers a DAG by testing the conditional independence of random variables. This method is based on conditional dependency among random variables, which is an extension of Pearl's work on Bayesian networks and the Inductive Causation Algorithm proposed in Pearl (1991). For an overview of the constraint-based learning method, please refer to Koller and Friedman (2009) or Scutari and Denis (2014). There are two steps in this method; the first one is the conditional independence test, and the second one is the edge orientation method. In addition, there are some methods, such as the Max-Min Hill-Climbing (MMHC) algorithm, that combine both of these approaches (Tsamardinos et al., 2006).

Our approach is based on the most fundamental and classic algorithm in the constraint-based learning method; it is a PC algorithm, named for its authors, Peter Spirtes and Clark Glymour, in Spirtes et al. (2000). This algorithm is used to recover a CPDAG when we are free of hidden and selection variables. Starting from a complete graph, in which each node connects with the rest, the PC algorithm gradually removes edges between nodes through a statistical independent test. The algorithm is based on marginally independent tests and then conditional on one vertex's performing conditional independent tests to construct the skeleton and so on. The direction is then added by the algorithm's identifying v-structure and further rules for directions. Kalisch and Bühlmann (2007) have proved the uniform consistency property of the PC algorithm in a high-dimensional setting when the number of variables is a polynomial of the sample size.

The PC algorithm does not work with a MAG or PAG due to hidden and selection variables. To overcome this limitation, an FCI algorithm (Spirtes et al., 2000), which is an improvement of the PC algorithm, is proposed. This algorithm, in addition to the PC algorithm (first-time orientation), incorporates additional steps to remove edges and reorients the graphs based on the PC-oriented collider structure graph. Specifically, the first two steps of the FCI algorithm are almost the same as those of the PC algorithm. In the following two steps, instead of the algorithm's checking all the subsets of the remaining random variables or d-separate set, a superset called *Possible-D-SEP*, as defined Spirtes et al. (2000), can be computed easily. For *G* as a mixed graph, Possible-D-SEP (X_i, X_j) in G is defined as: $X_k \in$ Possible-D-SEP (X_i, X_j) if and only if there is a path *p* between X_i and X_k such that, for every sub-path $< X_m, X_l, X_h >$ of *p*, X_l is a collider on the sub-path in *G*, or $< X_m, X_l, X_h >$ is a triangle of G. It can be shown that the first two steps of the FCI algorithm (or PC algorithm) generate sufficient information to compute a Possible-D-SEP set. Based on the Possible-D-SEP set, the FCI algorithm tests the conditional independence again and reorients the graph based on an updated skeleton and information on the separation set. In the final step, the algorithm uses the orientation rules described in Zhang (2008) to finalize the graph construction. The FCI algorithm has been shown to have the theoretical guarantee that, under some mild assumptions, the sample version of the FCI algorithm is consistent under the high-dimensional sparse setting (Zhang, 2008).

The learning with Possible-D-SEP sets is computationally demanding, rendering infeasibility when the size of the sets is larger than 25 (Colombo et al., 2012). To overcome this issue, some variants of the FCI algorithm, such as the RFCI algorithm and Conservative-FCI (CFCI) algorithm (Colombo et al., 2012), are proposed to help with large dimensional data. The motivation for using the RFCI algorithm is mainly that it tests a smaller number of variables for conditional independent. As a result, the presence of an edge in RFCI-PAG (PAG estimated by RFCI) has a weaker meaning than that of FCI-PAG (PAG estimated by FCI), and RFCI-PAG is theoretically a super-graph of FCI-PAG. RFCI, however, shows great computational advantage, with tolerable errors, when the dimensions of our data are high.

The CFCI algorithm is similar to the Conservative PC algorithm (CPC) proposed by Ramsey et al. (2012). This algorithm is based on two weaker conditions, "Adjacency-Faithfulness" and "Orientation-Faithfulness," in contrast to Markov and faithfulness conditions. The algorithm can potentially solve some situations when the transitive cause fails. As noted in Ramsey et al. (2012), however, CPC may not be as informative as the PC algorithm, implying that it might be too conservative to discover information. In fact, there is no complete step for orientation on the "unfaithful" mark. In addition, there is no theoretical superiority to assuming the orientation-faithfulness condition and no theoretical property of the further relaxation in CFCI, given that a PAG already assumes a less restrictive condition. Thus, we use FCI to learn a PAG, or RFCI when large dimensions lead to infeasibility or invalidity of FCI-PAG.

2.2.2 Stage 2: Estimating Causal Effects / Learn the Parameter

In the second stage, we estimate the scale of causal effects. This step is equivalent to conducting parameter learning of Bayesian networks in the language of artificial intelligence. Given an estimated graphical causal structure, the intuition when estimating causal effects is to control those non-causal effects, e.g., confounders, to adjust the estimated association to be consistent with causal effects. This adjustment is implemented by covariate adjustment.

The classic approach for covariate adjustment in the context of a DAG is the back-door criterion proposed by Pearl (1993). Specifically, a set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a DAG if:

1. None of vertices in Z is a descendant of X;
2. Z blocks every path between X and Y that has an arrowhead to X.

If Z satisfies the back-door criterion for a DAG G, we could use it to estimate the causal effect between X and Y in a DAG.

It is a sufficient condition to find a set of variables that adjust causal effects consistently. The back-door criterion is applicable, however, only when there is no hidden or selection variables. Because our context has hidden variables, it is infeasible to apply the classic back-door criterion. Therefore, a more generalized criterion is needed to estimate causal effects in a PAG.

We apply two recently developed generalized criteria to estimate causal effects. Worth noticing is that these criteria are available when there is no selection variable, which is satisfied by our first-stage results. The first criterion is a *generalized back-door criterion* (GBC) proposed by Maathuis et al. (2015). It generalizes the back-door criterion to the concept of *visible edge* introduced by Zhang (2008) as: given a MAG *M* / PAG *P*, a directed edge $X \rightarrow Y$ in *M* / *P* is visible if there is a vertex *Z* not adjacent to *Y*, such that there is an edge between *Z* and *X* that is into *X*, or there is a collider path between *Z* and *X* that is into *X*, and every non-endpoint vertex on the path is a parent of *Y*. Otherwise $X \rightarrow Y$ is said to be *invisible*.

Visible edges refer to situations in which there cannot be such a hidden confounder between *X* and *Y*. With the identification of a visible edge, one can extend the definition of a back-door path from *X* to *Y* in a PAG / MAG as a path between *X* and *Y* that does not have a visible edge out of *X*. Particularly in a PAG, it means a path that starts with $X \leftarrow *$, $X \circ - *$, or an invisible edge $X \rightarrow$. Zhang (2008) introduces two more definitions to completely define the GBC. One is a *definite non-collider*, which reduces to a non-collider in a DAG or MAG, but, in a PAG, it rules out the possible circle marks. A *definite status path* refers to a path in a partial mixed graph with all non-endpoint vertices as either a collider or a definite non-collider. Following this definition, all paths in a DAG or MAG must be definite status paths.

The definition of the CBC by Maathuis et al. (2015) is as follows: Let *X*, *Y*, and *Z* be pairwise disjoint sets of vertices in *G*. Then *Z* satisfies the GBC relative to ordered (*X*, *Y*) if the following two conditions hold:

1. *Z* does not contain possible descendants of *X* in *G*;
2. For every vertex $x \in X$, the remaining set of $Z \cup X$ blocks every definite status back-door path from x to any element of *Y* in *G*.

The back-door and GBC criteria are equivalent under the DAG framework for a single-intervention setting. Maathuis et al. (2015) propose a sufficient and necessary condition to find such a set that satisfies the GBC criterion. Because the condition requires a lot of graph knowledge, we do not present the condition here. However, we want to highlight that one could easily find the covariates for adjustment conveniently and feasibly compute the causal effects in the data analysis.

The GBC is a sufficient but unnecessary condition for estimating causal effects. Perkovic et al. (2015) further propose a complete GAC that is necessary and sufficient for all of the four types of diagrams that we discuss. The GAC is based on the concept of *amenability*: If a graph *G* is *adjustment amenable* relative to (*X*, *Y*), then every possibly directed proper path from *X* to *Y* in *G* starts with a visible edge out of *X*. This concept is similar to the definition of the back-door path, but it is defined only on a possibly directed proper path, which relaxes the requirement of a directed path to that of no arrowhead as pointing to the starting vertex. In addition, a path is *proper* from Set *X* to Set *Y* if its first node is in *X*.

The definition of the GAC given by Perkovic et al. (2015) is as follows: Z satisfies generalized adjustment criterion relative to (*X*, *Y*) if:

1. *G* is an adjustment amenable relative to (*X*, *Y*);
2. No element in *Z* is a possible descendant in *G* of any *W*, except *X*, which lies on a proper possible directed path from *X* to *Y*;
3. All proper definite status non-directed paths in *G* from *X* to *Y* are blocked by *Z*.

It is straightforward that both the GBC and GAC are based on intuition in regard to blocking non-causal paths by conditioning on covariate adjustment. Even though the GAC compensates for the shortcomings of the GBC, as it provides only a sufficient condition for an adjustment set, while the GAC provides a necessary and sufficient condition, the GAC does not provide an easily checkable condition, and, thus, there is no algorithm-perspective construction of an adjustment set based on GAC.

Having covariate adjustment set *Z* via the GBC and/or GAC, one can estimate the causal effects in a PAG. These effects are attained by the definition of the adjustment criterion whereby the motivation of the GBC or GAC is: the set of variables *Z* of *G* satisfies the adjustment criterion relative to (*X*, *Y*) if, for any probability density f compatible with G, we have:

$$f(y \mid do(x)) = \begin{cases} f(y \mid x) & \text{if } Z=\varnothing \\ \int_z f(y \mid z, x) f(z) dz = E_z\{f(y \mid z, x)\} & \text{otherwise} \end{cases} \tag{2}$$

Here, the "do" operator refers to the intervention operator proposed by Pearl (1995) for calculating causal effects in non-parametric models based on the intervention. Equation (2) ensures the identifiability of the estimate of the causal effect between variables by transforming intervention probability into conditional probability so that we can estimate the causal effect based on observational study. Once the adjustment set is found, under the Gaussian distribution assumption, the mean of the causal effect is equivalent to:

$$\frac{\partial}{\partial x} E[Y \mid do(X = x)] \ , \tag{3}$$

$$\text{that is } \frac{\partial}{\partial x} E[Y \mid X = x, Z = z] , \tag{4}$$

Note that we focus only on the linear causal effect. The formula above simply reduces to an econometric model shown as:

$$y_i = \beta_1' X_i + \beta_2' Z_i + c + \varepsilon_i , \tag{5}$$

where y_i represents a single vertex that is causally affected, X_i represents a set of vertices that exerts causal effects, and Z_i represents vertices in the adjustment set. β_1' is the parameter vector that capture the causal effects of X_i, which is the one of interest that is to be estimated.

The reduced Model (5) has consistent interpretation in econometrics. The GAC and GBC suggest that controlling Z_i eliminates the non-causal effects of X_i on y_i, which is equivalent to taking Z_i as a control variable to alleviate confounding factors econometrically. Our approach, however, shows its advantage by pinpointing the correct control variables, instead of choosing them simply by assumptions.

3 Data

We use a unique dataset that records app usage behavior of 600 randomly sampled smartphone users in China. For each, we have one observation of the weekly frequency of clicking on all attainable apps on the main user-interfaces of their smartphones. We collect the data for one non-holiday week, starting February 7, 2015, for the purpose of model estimation.

To check the robustness of our finding with respect to stationarity over time, we additionally collect datasets in the same way but for the time windows of the next two weeks (the weeks of February 14, 2015, and February 21, 2015). Note that these two weeks cover the Spring Festival (Chinese New Year), which is an 11-days-long national holiday. This enables us to test whether the causal effects are, in general stationarity, between holiday and non-holiday times. In addition, to check sampling errors, we collect datasets for another sample of 600 individuals that has no overlap with the original sample, in the same way as we execute the original dataset for the same three weeks. In sum, we have two cross-sectional samples and three time periods for each.

The final data (including data for robustness check) included 1,122 different apps, of which 898 appear in the data for estimation. To help readers to have a better understanding of the app market in China, we list the top-50 frequently used apps in China and note the developer and alliance of each in Table 1. It is apparent that the app market is not fragmented, suggesting that major developers, such as Baidu, Alibaba, and Tencent (BAT), dominate the app market.

Table 1 App Number, App Name, and Corresponding Developer or Affiliation

App No.	App Name	Developer or Alliance	App No.	App Name	Developer or Alliance
1	WeChat	T***	40	91 Lotto	B
2	T Map	T	41	JD.com	T
3	QQ	T	42	B Search	B
4	T Video	T	46	Ali Pay	A
5	QQ Space	T	48	Wo Music	O
6	Weibo	A*	54	MeiTuan	A
7	Other QQ Product	T	55	B Map	B
9	Voice Control	O	59	Moji Weather	O
10	Didi	T	60	QQ Music	T
13	T News	T	68	Iqiyi	O
14	Sogou Typing	S****	72	Tieba	B
15	QQ Browser	T	74	B Wenku	B
17	Youku Video	A	87	Xunfei Plugin	O
18	Kugou Music	O******	88	Baidu Assistant	B
20	Gaode Map	A	91	ZD Clock HD	O
21	B Category	B**	101	Wangyi News	Y*****
22	UC Browser	A	109	WIFI	O
23	360 Guide	O	132	Sohu News	S
24	TouTiao	O	146	App Store	O
27	Android MKT	O	149	Sohu Video	S
29	MiLiao	O	152	Fun TV	O
32	91Phone Assistant	B	188	App Market	O
33	B Map Plugin	B	196	Coolpad Weather	O
35	Sina News	A	239	Kowo Music	B
39	Taobao	A	332	Momo	A
*A = Alibaba **B = Baidu ***T = Tencent			****S = Sohu *****Y = Wangyi ******O = other or independent developers		

In the data for estimation, the weekly average clicking rates for different apps exhibit a typical long tail, with WeChat's on the very left-hand side, as shown in Figure 1 (a). In Figure 1 (b), a closer examination of the top-50 frequently used apps listed in Table 1 shows that the usage of WeChat (the very left-hand side) is at least two times that of the second most frequently used app, confirming its mega status in app usage. We further check the stationarity by including the dataset for a robustness check and depict the average weekly clicking rates across 1,200 individuals over three weeks in Figures 1 (c) and (d). A comparison with Figures 1 (a) and (b) shows similar shapes but fatter tails for their distributions.

Figure 1 App Weekly Average Clicking Rates of Estimation Sample and Pooled Sample

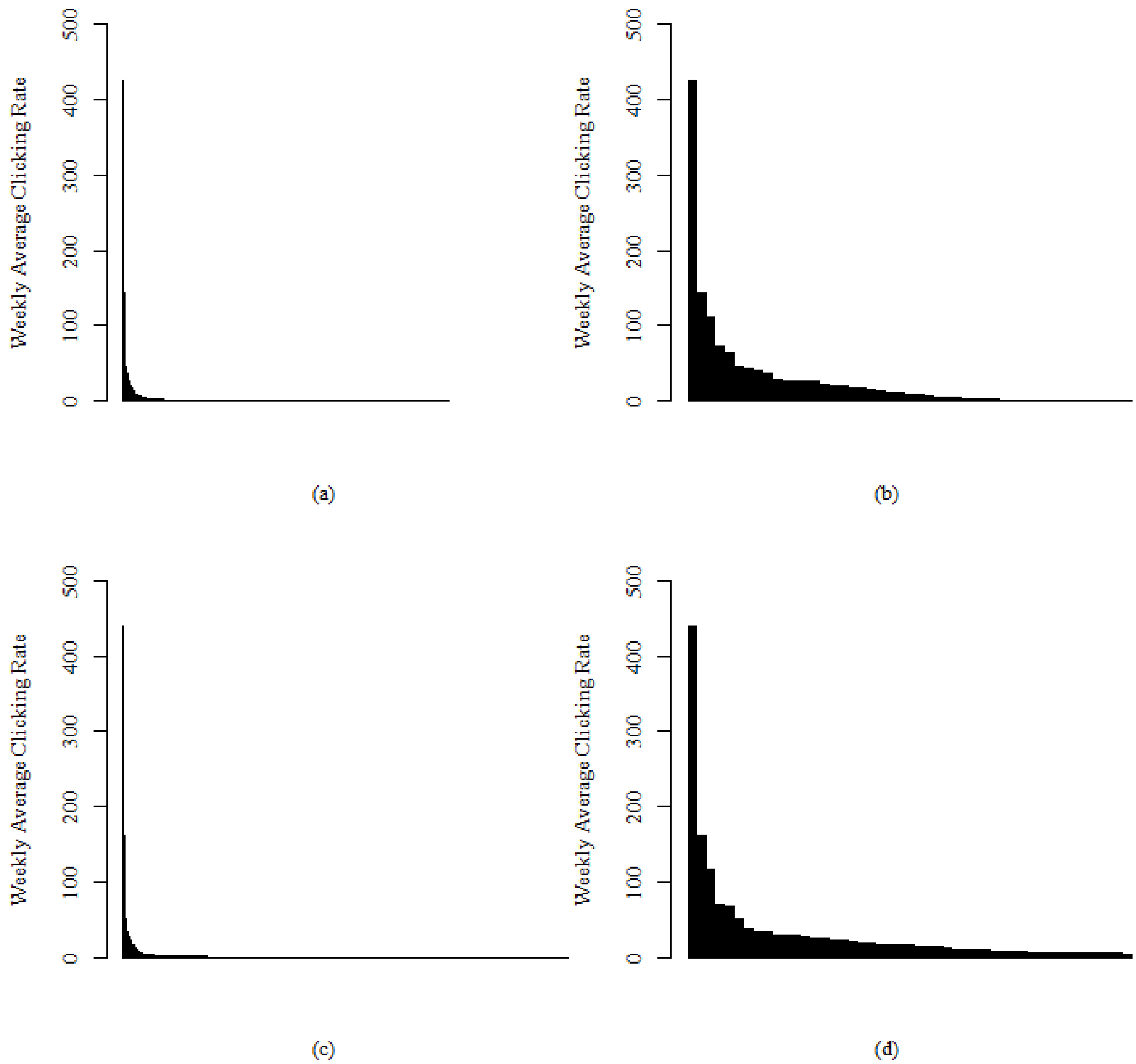
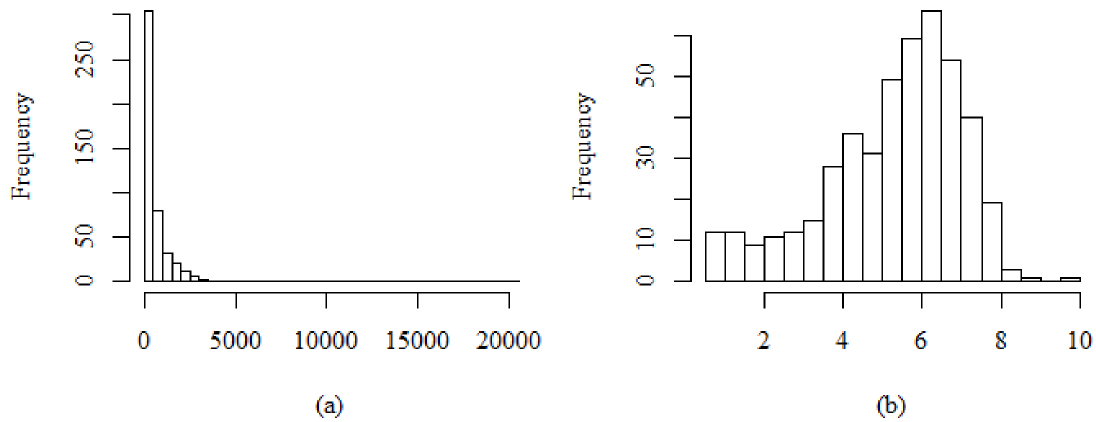


Figure 2 (a) presents the distribution of WeChat usage, for which the clicking rates are quite skewed, with the majority of clicking rates as less than 5,000, with the maximum above 20,000. The skewness suggests a potential problem if we want to make use of a multivariate normal distribution for the estimation of causal effects. Therefore, we take a log transformation of our data to approximate a multivariate normal distribution. For WeChat, the transformed data are shown in Figure 2 (b).

Figure 2 Distribution of WeChat Usage



4 Estimation Results

Our goal is to capture the causal relationships between different apps, and if there is such a relationship, we hope to estimate the causal effects based on the observational data. We assume that there is (possibly) no directed cyclic graph between apps, which is practical in reality and satisfies the faithfulness condition. Considering that there might be hidden apps behind the data and that selection bias may exist, instead of constructing a CPDAG, we use a PAG to model our data to reduce bias and attain lower variance than would be seen in a CPDAG. In addition, the space of PAGs is smaller than that of CPDAGs, which makes the search more feasible. When the sample is large, the same data with a single PAG can solve a lot of meaningful questions behind the app data, while a CPDAG might give us a different graph structure. Given an estimated PAG, in the second stage, we further quantitatively estimate the causal effects by applying the GAC and GBC to find the valid adjustment set.

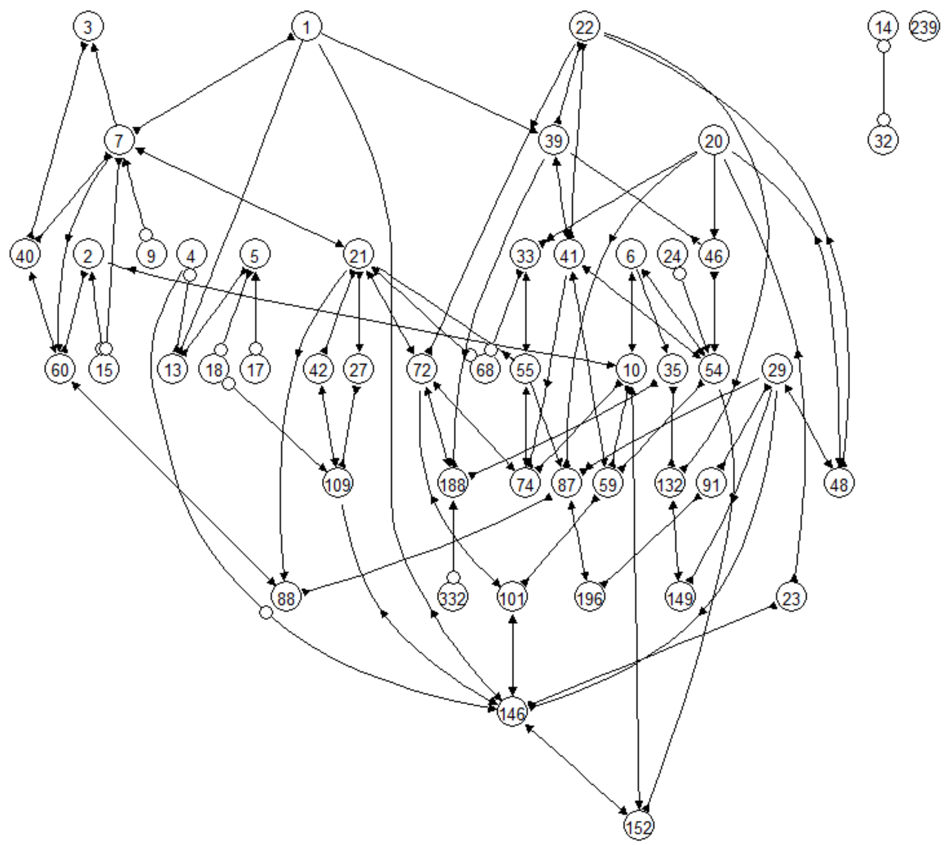
We estimate the causal relationship of the top-50 most-used apps only in the main model for following reasons. First, the usage of many rarely used apps exhibits no dependency on the rest. Having a smaller set generates a more concise presentation. Second, those rare apps typically focus on niche markets, which have a less significant impact on the app market as compared to that of top ranked apps. Third, methodologically, (log transformation of) usage of rarely used apps can barely satisfy normal distribution assumptions, which could not only lead to problematic results but also contaminate the results of those frequently used apps. To alleviate concerns about this approach, we extend the set to include more apps for the analysis in Section 5. Compared to a PAG estimated with an extended set of vertices that includes more apps, PAG estimated with top 50 apps shows that the spillover effects of our focal app, WeChat, are well captured and depicted locally.

We present the results as follows. First, we provide the causal structure of app usage graphically as the PAG that we determined through the FCI algorithm. Second, we measure the spillover effects quantitatively based on the estimated PAG using the GAC and GBC criteria, with econometric interpretation. The quantitative measurement provides further information on the causal effect as positive or negative as well as its strength. Third, to show the value of the graphical model for estimating causal effects, we extend our discussion to cases that are assumed to be estimated without knowing the causal structure from a PAG or with an incorrect adjustment. In those examples, the existence of spillover effects is ruled out by graphical results and interpretation; however, these effects are estimated to be significantly not zero due to the bias of incorrect adjustment.

4.1 Stage 1: Graphical Results

We present our estimated causal diagram in Figure 3. In this diagram, each node shown as a number represents an index of one specific type of app, which is the App Number in Table 1. The diagram explicitly displays local causal effects of WeChat (App 1). Note that the edges out of WeChat are visible ($1 \rightarrow 13$ and $1 \rightarrow 39$). This indicates that there are no unobserved confounders behind a direct edge and that each directed edge out of WeChat represents corresponding causal effects explicitly. Specifically, the diagram shows that WeChat has direct spillover effects on two apps: Tencent News (App 13), a news app developed by the same parent company, and Taobao (App 39), the leading shopping platform in China, developed by the Alibaba group. Other than these two apps, WeChat exhibits direct correlations with other QQ products (App 7) and Appstore (App 146), driven by unobserved confounders (as they are connected bi-directly). Figure 3 suggests that the correlation between all other apps and WeChat is confounded by hidden variable(s) that are not observed and/or conditionally driven by colliders (observed selection variables) in the data. In sum, the diagram suggests that, even though WeChat dominates smartphone user app use, its direct externality toward other apps is not as strong as we had expected. In fact, it is so limited that only two other apps are affected directly.

Figure 3 PAG of (Top 50) App Usage Causal Structure



The finding suggests that, although associations between WeChat and other focal apps might be found, they are not necessarily explained causally. In fact, for the majority, it is confounders rather than spillover effects from WeChat that explain the association. App developers should be cautious about being deceived by associations when analyzing attribution and collaboration, as the identities of factors that determine the usage of apps might not be the same ones that show the association of usage with the focal app. Given that a connection to such mega apps might incur high costs, our approach provides a tool that allows app developers to visually and directly examine the spillover effects from WeChat and other apps. Our approach provides an understanding that is deeper than that provided by superficial association and helps app developers with decision making with regard to developing collaborations and connections for economic interests.

Worth noticing is that the estimated PAG contributes not only to qualitative but also to quantitative findings. Any node without a (possible) causal path from WeChat is indicated as having no causal effects from WeChat. Therefore, it can be concluded quantitatively that all nodes in Figure 3, other than Tencent News and Taobao, receive zero causal effects from WeChat.

4.2 Stage 2: Quantitative Results

Given the results for apps that receive zero causal effects from WeChat, however, for apps that receive non-zero spillover effects, we need to estimate the scale of them quantitatively in additional steps. Specifically, to avoid potential biasness due to observed confounders, unobserved confounders, and selection variables, we use the causal structure estimated by the FCI algorithm in Figure 3 to adjust non-causal factors, following the GAC and GBC. Figure 3 shows that non-causal paths are all blocked by colliders for both Tencent News and Taobao, implying that the adjustment set Z is an empty set, following the GAC or GBC. The model simply reduces to a linear regression with the usage of the focal app, WeChat, as the only independent variable.

Table 2 shows that the spillover effects of WeChat are positive for both Tencent News and Taobao. Specifically, for an average user of WeChat, a 10% increment of usage of WeChat leads to 7.25% additional usage of Tencent News and 8.33% more usage of Taobao. This suggests that, as different types of apps are created by the same developer, the functionality of WeChat complements that of Tencent News effectively. WeChat users who are interested in reading news are successfully directed to the news app developed by the same company, indicating one more step to the goal of full service of Tencent. However, the spillover effect on Taobao suggests positive externality to Alibaba, the major competitor of Tencent, given that Tencent has its own online shopping platform and other ecommerce platforms as a strategic alliance. The existence of spillover effects suggests a loss of users with the intention of online shopping, as provided by the competitor.

Table 2 Estimation Results

Parameter	Tencent News (13)	Taobao (39)
β_1	0.35***(0.02)	0.40***(0.03)
c	0.20*(0.10)	0.35*(0.14)
Marginal Effects (10% in X)	7.25%	8.33%
Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05	

4.3 Estimates based on Incorrect Adjustments

The value of a graphical model is not limited to aiding the estimation of causal effects, as shown in Section 4.2. Moreover, the estimated causal structure itself encodes enormous interpretable information on causal effects that helps researchers to have an understanding of correctly adjusted causal effects, which would otherwise be incorrectly estimated. In this section, we present several common representative cases in econometric causal inference that appear in our context, including unadjustable latent confounding bias, adjustable latent confounding bias, and endogenous selection. Note that the value of a PAG is not limited to the three cases that we mentioned above. In addition, it can solve over-controlled bias, observed confounding bias, and so on (Elwert, 2013). We skip those issues, however, because those cases do not appear in our context. Further, an incorrect adjustment can happen in any vertices in our data. Due to space limitations, we illustrate only three cases that occur in our data through three representative vertices.

4.3.1 Unadjustable Latent Confounding Bias

Based on the interpretation rule of a PAG, a bi-directed edge $A \leftrightarrow B$ suggests that A has no causal effects on B (due to the arrowhead at A), and B has no causal effect on A (due to the arrowhead at B). There is no ancestral relationship between A and B, but they are adjacent. Therefore, the association between A and B can be explained only by latent confounder(s) (Kalisch et al. 2012). Because the confounder(s) are unobserved, the confounding bias cannot be adjusted. Therefore, a linear regression model cannot correctly estimate the causal effect between A and B. A naïve regression of A on B would induce the confounding bias due to the unobserved confounder.

In our example, unadjustable latent confounding bias exists between the usage of WeChat and that of other QQ products as well as between the usage of WeChat and that of Appstore. The interpretation of a PAG suggests no causal relationship between WeChat and QQ products or Appstore. However, researchers would estimate the causal effect as positively significant if they have no information about the causal structure and mistakenly regard the association as causal effects. We estimate the association and compare it with the causal effect based on a PAG in Table 3.

Table 3 Example of Unadjustable Latent Confounding Bias

	Parameter	Other QQ product (7)	App store (146)
Association	β_1	0.45***(0.03)	0.17***(0.02)
	c	1.25***(0.13)	0.02(0.01)
Causal Effects by PAG		0	0
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05			

This result shows the methodological advantage of a PAG for estimating causal effects from observational data with hidden confounder(s). Other methods for causal inference alleviate the confounding bias by controlling potential confounding factors, such as propensity score matching. However, such an approach is limited to conditioning on observed confounder(s) only, leading to biased estimation when unobserved confounders exist. The PAG approach, in contrast, infers the existence of an unobserved confounder, which further helps researchers to adjust causal effects correctly.

4.3.2 Adjustable Latent Confounding Bias

Latent confounding variables are adjustable when observed intermediate non-collider vertices exist on the causal path from the latent confounder to focal variables. The simplest example is $A \leftrightarrow B \rightarrow C$. In this example, *A* has no causal effect on *B* and *C*. However, *A* and *C* show an association due to a common confounder between *A* and *B*. This confounder exhibits a causal effect on *C* indirectly through *B*. Given that the edge between *B* and *C* is visible because A points to B and B is a non-collider, conditioning on *B* would control the causal effects from the latent confounder to *C*. Therefore, a linear regression of *B* and *C* would adjust the latent confounding bias. If $A \leftrightarrow B \rightarrow C$ is the only unblocked path between *A* and *C*, the regression that suggests zero as the coefficients for *A* can be used as the validation for the bias of the adjustable latent confounding variables.

In our example, one apparent path with latent confounding bias is from WeChat to QQ (App 3), another instant messaging app developed earlier by Tencent, through other QQ products, shown as $1 \leftrightarrow 7 \rightarrow 3$. Note that there is no other unblocked path between WeChat and QQ. The graph suggests that adding usage of other QQ products in an adjustment set *Z* would control the causal effect from WeChat to QQ. The results in Table 4 confirm our expectation by showing the causal effect of App 1 on App 3 to be insignificantly different from 0. The estimation of causal effects without controlling the usage of other QQ products would result in a biased estimation due to failing to adjust for the effect of unobserved confounder(s) between App 1 and App 7.

Table 4 Example of Adjustable Latent Confounding Bias

	Parameter	Adjusted	Unadjusted
Association	β_1	0.01(0.02)	0.45***(0.03)
	β_2	0.96***(0.02)	
	c	0.29***(0.07)	1.49***(0.01)
Causal Effects by PAG		(7) has effect on (3)	0
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05			

This finding also shows considerable consistency with recent observations and anecdotes about the relationship between QQ and WeChat, two instant messaging apps by the same developer, from an industry perspective. An industry observer reported that WeChat was designed strategically to differentiate itself from QQ, such that very limited substitution exists (Geekpark, 2013). This observation was confirmed by the CEO of Tencent (ithome, 2013). Individual users would be driven to use these two apps based on different functional needs, such that no direct dependency between these two apps should exist. Other confounder(s), however, might encourage usage of both apps, which would result in association, consistent with our estimation results.

4.3.3 Control Over Selection Variable

In the two cases above, we show the potential bias due to failing to control non-causal factors. In econometrics, such cases are typically due to failing to have confounders as valid control variables. This leads to a concern about whether this means that we should have as many control variables as possible to alleviate biasness to the maximal level. In this section, we present a problematic estimation if the control variable is a collider (selection variable), rather than a confounder, on the path. Note that the PAG identifies the role of each node on a path as a collider (or not). This again shows a methodological advantage as compared with models that have an uncertain status of the confounder or collider of each control variable before estimation.

The problem of endogenous selection bias occurs when a collider is added into the adjustment set Z. Specifically, conditioning on the common outcome of two variables induces a spurious association between them for at least one value of the collider (Elwert 2013). For example, when we have a PAG shown as $A \leftrightarrow B \leftrightarrow C$, this suggests one possible structure with two latent variables, revealed as $A \leftarrow L_1 \rightarrow B \leftarrow L_2 \rightarrow C$, and that *A* does not have any causal effect on *C* if there is no other path or if all other paths are blocked. However, if we condition on observed vertex *B*, the causal structure will be replaced as $A \leftarrow L_1 - L_2 \rightarrow C$, where *A* is associated with *B* due to the spurious path between L_1 and L_2 . Because L_1 and L_2 are unobservable, and thus cannot be added into the adjustment set to block this spurious path, a spurious causal effect will be estimated to represent the endogenous selection bias.

Table 5 Example of Endogenous Selection Bias

	Parameter	Adjusted
Association	β_1	-0.07**(0.02)
	β_2	0.62***(0.03)
	c	-0.23*(0.10)
Causal Effects by PAG		0
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05		

There are many potential examples of endogenous bias if we do not design the adjustment set in the correct way. We take a causal relationship between WeChat and 91 Lotto (App 60), the leading online lotto marketplace in China, as an example. According to the estimated PAG, the causal effect from WeChat to 91 Lotto is 0 because there is no causal path from WeChat to 91 Lotto. However, if we erroneously add usage of other QQ products (App 7) into the adjustment set *Z*, the causal effect from WeChat to 91 Lotto is estimated to be significantly negative, as shown in Table 5. This is because conditioning on other QQ products opens a spurious confounding path between WeChat and 91 Lotto, whose confounder is unadjustable ($1 \leftarrow L_1 - L_2 \rightarrow 60$). This example provides important information for researchers: Adding an incorrect control variable risks deteriorating the estimation of causal inference.

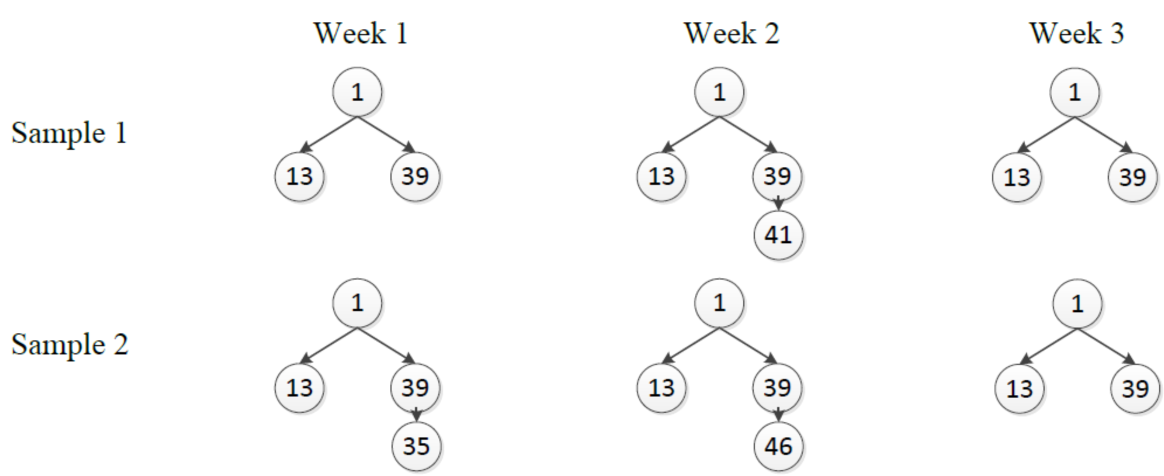
5 Robustness Checks

As presented in this section, we conduct a robustness check to ensure the consistency of the findings and to eliminate potential explanations, such as sampling errors and time-specific factors. Given the nature of the two-stage estimation for causal effect estimation, we first check the consistency of graphical outputs, as discussed in Section 5.1, and then check that of quantitative results, as presented in Section 5.2.

5.1 Check Graphical Results

To eliminate concern about sampling errors, we use the alternative sample, in which there are no overlapping individual users. To eliminate the concern about time-specific factors, we collect further data for the next two weeks. Note that two weeks after the time of the original observation is a national holiday. It would imply a high degree of consistency if the spillover effects of WeChat in the original PAG are the same or close to that in the PAG of the holiday. Given the two sets of samples and the three time periods for each, we could estimate six PAGs. For succinct presentation, we draw graphs of only the causal paths from WeChat. Further, we apply RFCI when FCI is infeasible or invalid. The PAGs are displayed in Figure 4**.**

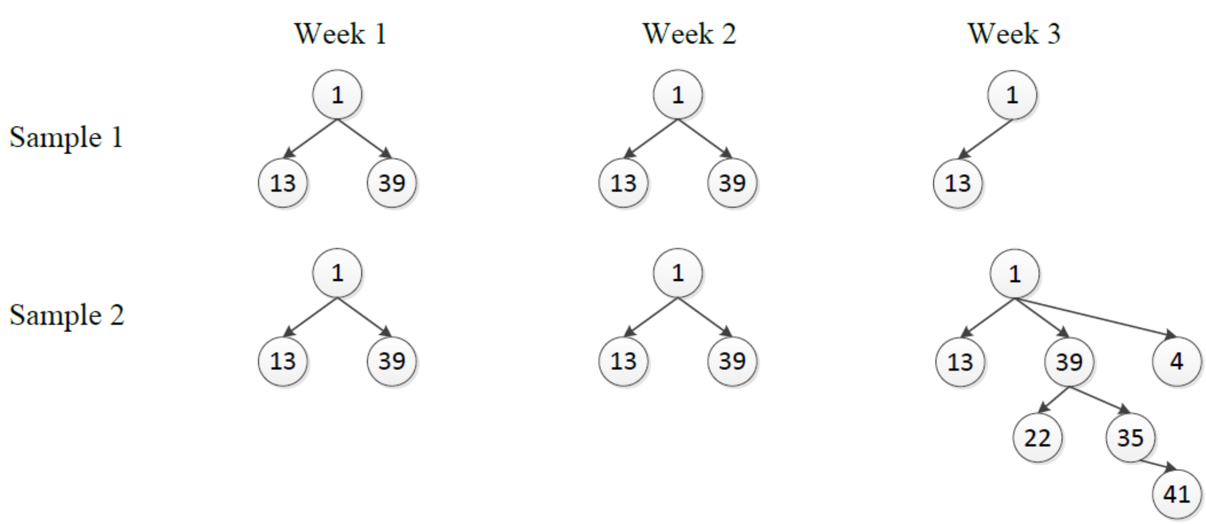
Figure 4 PAGs for Two Sets of Individuals and Three Time Periods with Alpha = 0.01



As can be seen in Figure 4, the causal paths from WeChat are quite consistent for all six samples. All PAGs show direct causal effects on Taobao (App 13) and Tencent News (App 39), suggesting that the causal effects identified in the original sample are robust to different samples and, thus, robust to sampling errors and time-specific factors. The mild discrepancy lies in the PAG of Week 2 and Sample 1, which exhibits an indirect causal effect on App 39 through App 41; the PAG of Week 1 and Sample 2 exhibits an indirect causal effect on App 35 through App 39; and the PAG of Week 2 and Sample 2 exhibits an indirect causal effect on App 39 through App 46. These effects are quite unstable, however, and could be attributed to sampling errors or time-specific factors.

The PAGs are drawn based on conditional independence tests with a threshold for significance fixed at a certain level (alpha) to control type-1 errors in the statistical hypothesis testing framework. The level of alpha could be regarded as a trade-off between the probability of having an error in independence and the power of detecting dependence. As a result, PAGs estimated on the same observation but with different levels of alpha might exhibit different patterns. As seen in Figure 5, to examine the impact of alpha, we relax the alpha from 0.01 to 0.05 and redraw PAGs in the same way as in Figure 4.

Figure 5 PAGs for Two Sets of Individuals and Three Time Periods with Alpha = 0.05



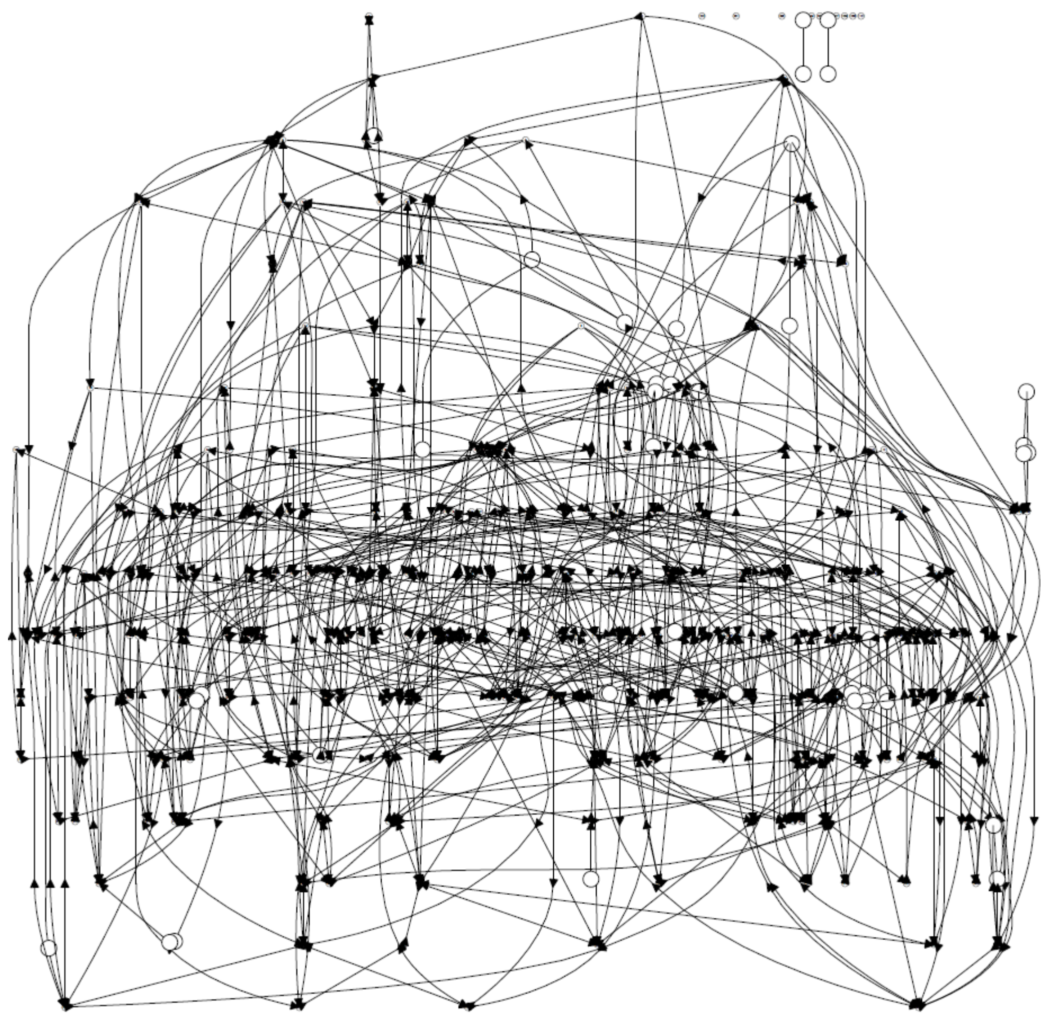
This figure also shows a high level of consistency of the causal structure. The majority of the graphs show a direct causal effect on Taobao (App 13) and Tencent News (App 39). The graph of Sample 1 in Week 3 does not have a causal path on Tencent News. Instead, it exhibits a bi-directed edge between WeChat and Tencent News. In addition, the graph of Sample 2 in Week 3 shows additional causal paths, including a direct causal effect on App 4. This is not surprising, however, because, as we increase the level alpha, we will have more vertices connected because the power of detecting the dependency signal increases.

Figure 6 PAG of Top-100 Popular Apps

The third robustness test is for the set of apps that we use for estimation. Additional information of app usage would provide more information on causal relationship identification.

Therefore, robust causal relationships should stay constant if we increase the size of the vertices set. Specifically, we estimate two more PAGs with the top-100 frequently used apps and top-300 frequently used apps, correspondingly shown in Figures 6 and 7, with the alpha as fixed at 0.01. The estimation for the PAG with the top-300 apps is implemented with the RFCI algorithm due to the infeasibility of applying the FCI to high dimensional data.

Figure 7 PAG of Top-300 Popular Apps



Due to the large scale of the vertices, the readability of the graph can be difficult. We examine the adjacent matrix and find the existence of causal paths from both WeChat to Taobao and to Tencent News, as seen in Figures 4 and 5. Specifically, the PAG of the top-100 apps shows causal paths from WeChat to Taobao and to Tencent News as the only causal paths, which is exactly the same as seen in the PAGs of the top-50 apps. Further the PAG of the top-300 apps has causal paths to Taobao and to Tencent News as the only two direct causal paths. These consistencies suggest that our original model for the top-50 apps is able to capture most of spillover effects of WeChat. The PAG of the top-300 apps, however, has additional indirect causal paths to two apps, one of which is not included in the PAG of either the top-50 apps or of the top-100 apps. However, we reserve a conservative attitude toward these two causal paths for the following two reasons: (1) For the usage distribution of those less popular apps (of the top-100 popular apps set), it might be difficult to approximate the Gaussian distribution even after logarithm transformation. As we noted, when we took the logarithm of the app usage, if there was a great deal of zero usage, it could cause enormous skewness; and (2) RFCI-PAG is recognized as a super-graph of FCI and has weaker meaning in regard to the presence of edges than does the FCI, as shown in Colombo et al. (2012). Both reasons cast doubt on the robustness of these two causal effects.

5.2 Check Quantitative Results

As discussed in this section, we conduct a robustness check for the scale of causal effects. Specifically, we estimate causal effects from the data of distinct samples and time periods. Note that the estimation is based on the learned structure in the graphical results, and given a PAG, the specification for learning the graph has no impact on the quantitative estimation results. Therefore, there is no need to investigate the robustness of the alpha level or size of the vertices.

We first estimate spillover effects of WeChat on Tencent News and Taobao with distinct samples across different time periods separately, using the main model (5). The estimation results are shown in Table 6. Our results suggest a high degree of consistency across distinct samples and time periods. In all specifications of samples, the spillover effects on both Tencent News and Taobao are estimated to be positive, with the effect on Taobao as stronger quantitatively. The scales of effects are quite close among all six samples. The consistency of results based on different samples proves the robustness of our quantitative estimation.

Table 6 Comparing Quantitative Results Separate Samples

		Week1		Week2		Week3	
		Tencent News (13)	Taobao (39)	Tencent News (13)	Taobao (39)	Tencent News (13)	Taobao (39)
Sample 1	β_1	0.35*** (0.02)	0.40*** (0.03)	0.32*** (0.12)	0.37*** (0.02)	0.32*** (0.02)	0.33*** (0.03)
	c	0.20* (0.10)	0.35* (0.14)	0.12 (0.08)	0.09 (0.11)	0.18* (0.08)	0.24* (0.12)
Sample 2	β_1	0.38*** (0.02)	0.39*** (0.03)	0.36*** (0.02)	0.33*** (0.03)	0.35*** (0.02)	0.35*** (0.03)
	c	0.25* (0.10)	0.35* (0.15)	0.09 (0.09)	0.23 (0.12)	0.14 (0.08)	0.28* (0.11)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

Finally, note that pooling those six samples generates a sample of 1,200 individual smartphone users with repeated measures longitudinally. This pooled sample provides us with the opportunity to tease out individual-specific factors and time-specific factors to alleviate confounding bias. Note that our model suggests that no confounder exists on the causal paths from WeChat to Tencent News and to Taobao. Therefore, we expect estimates of parameters in a model with controlled individual-specific factors and time-specific factors to be similar to the estimates in former specifications. We control individual-specific factors and time-specific factors by adding fixed effects and specify the model as follows:

$$y_{it} = \beta_1' X_{it} + \beta_2' Z_{it} + c + \xi_i + \tau_t + \varepsilon_{it} \tag{6}$$

where ε_{it} are unobserved error terms following a Gaussian distribution. ξ_i and τ_t capture individual-specific unobserved effects and time-specific unobserved effects, respectively. Z_{it} is an empty set based on the GAC and GBC when estimating causal effects of WeChat on Tencent News and on Taobao. In addition, we estimate the causal effects by applying an OLS model without fixed effects on pooled data for comparison. We report the estimates in Table 7.

Table 7 Spillover Effects Based on Pooled Sample

	Parameter	Tencent News (13)		Taobao (39)	
		FE	Pooled OLS	FE	Pooled OLS
Association	β_1	0.33***	0.35***	0.32***	0.37***
		(0.01)	(0.01)	(0.02)	(0.01)
	c	-0.48*	0.15***	-1.05	0.25***
		(0.44)	(0.04)	(0.61)	(0.05)
	FE	Not Report	NA	Not Report	NA
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05					

As we expect, parameter estimates for causal effects in Model (6) are very close to those of the original model (5). This implies the non-existence of a confounder that is encoded in the graphical model and further supports the robustness of our quantitative results.

6 Conclusions, Limitations, and Future Research

The instant messaging app WeChat exhibits a mega status in the app market and exhibits dominance in terms of usage among Chinese smartphone users. However, its externality toward other types of apps has not received sufficient attention. Research is needed to investigate the spillover effects of such apps and to determine implications for the value that it creates for its developers as well as the value that it delivers to developers of other apps.

We combine a state-of-the-art machine learning method with an econometric approach to study the spillover effects of WeChat. Specifically, we apply an FCI-PAG method to determine the causal structure of app usage from observational data and estimate the spillover effects quantitatively based on the graphical outputs, the generalized back-door criterion, and generalized adjustment criterion. By applying our model to the app usage data of 600 Chinese smartphone users, we identify the set of apps that causally receive spillover effects from WeChat, the set of apps that shows association with WeChat due to observed or unobserved confounders, and the set of apps whose usage are independent of that of WeChat. We find that, counterintuitive to the belief of the industry, WeChat has quite limited external effects on the usage of other apps: among the top-50 and the top-100 apps, only two, Tencent News and Taobao, are shown to be causally positively affected by the usage of WeChat. Even when we extend the set to 300 apps, only these two apps receive spillover effects directly. The rest receive no causal effects from WeChat. To illustrate the importance of determining the causal structure and the value of quantitative information encoded in a graphical model, we further intentionally specify the econometric model with an incorrect adjustment set to show the erroneous estimation without the graphical results in the first stage.

Finally, we present the robustness of this approach by conducting a comparison of graphical estimates and quantitative estimates across samples in different time periods with different individuals. Using a pooled sample with repeated measure of individuals, we estimate the model with individual- and time-specific effects controlled to show the robustness of visible edges. In sum, this empirical study is the first to examine spillover effects of a mega app, such as WeChat. It provides researchers and app developers with a causal understanding, which is deeper than that provided with a superficial association explanation, and contributes to the analysis of attribution and decision-making about collaboration.

This paper is also the first to apply recent developments in machine learning-enabled causal inference models, such as FCI-PAG, plus a GAC and/or GBC estimation approach in business and economic research. Compared with past research methods, our approach relaxes the need for assumptions to identify causal effects with observational data but incurs a cost for obtaining additional information (hidden variables) when determining the causal structure. However, because data have become increasingly less expensive in this age of big data, this approach has the potential to be widely applied in estimating causal effects in business analytics research. Our work, as pioneering research that applies FCI-PAG plus GAC and/or GBC estimation, not only presents the spillover effects of WeChat but also shows a good fit of this advanced method in the context of business analytics research.

Our research is subject to limitations. These limitations typically relate to restrictions of the integration of the FCI-PAG-GAC or GBC approach and econometric methods, which, in turn, opens up avenues for future business analytics research. First, a more flexible model might be developed to allow for non-Gaussian distributed data. In our context, we use a log transformation to approximate our data to Gaussian. More complicated cases, such as ordinal choice data, however, might require a nonparametric graphical model to estimate. Second, even when we estimate a fixed-effects model with individual- and time-specific factors separately in the robustness check section, we notice that such factors cannot be added into the graphical estimation (first-stage estimation) due to the challenge of the independence test between Gaussian-distributed variables and dummy variables. A more generalized model that allows fixed effects might need to be developed to provide a more consistent (between graphical and quantitative models) and accurate estimation with repeated measurements graphically. Third, more graphical model-based econometric tools should be developed to help to estimate or validate the outputs from an FCI-PAG-GAC/GBC approach. For example, a searching algorithm for generalized conditional instrument variables that works in a DAG/CPDAG setting should be extended to an MAG/PAG setting to help to validate the visible edge.

These three recommendations, based on the limitations of our research, would help to further develop the connection between econometric and graphical models. Given the similarities of the nature of these two methods, more complete integration should be promising in future research. Further, our method can be easily adapted to other research contexts, such as online social networks and recommendation systems. Given the power of drawing causal inferences from observational data, we expect that more fruitful applications of this approach would contribute to a better analytical understanding of business.

7 References

Buntine, W., 1991, July. Theory refinement on Bayesian networks. In *Proceedings of the Seventh conference on Uncertainty in Artificial Intelligence* (pp. 52-60). Morgan Kaufmann Publishers Inc.

Campos, L.M.D., Gámez Martín, J.A. and Puerta Castellón, J.M., 2002. Learning Bayesian networks by ant colony optimisation: searching in two different spaces. *Mathware & soft computing. 2002 Vol. 9 Núm. 2 [-3]*.

Carare, O., 2012. The impact of bestseller rank on demand: Evidence from the app market. *International Economic Review*, *53*(3), pp.717-742.

Cooper, G.F. and Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, *9*(4), pp.309-347.

Chan, C. (2015, August 6). When One App rules them all: The case of WeChat and mobile in china. Retrieved August 15, 2016, from A16Z, http://a16z.com/2015/08/06/wechat-china-mobile-first/

Colombo, D., Maathuis, M.H., Kalisch, M. and Richardson, T.S., 2012. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp.294-321.

Cormack, M. (2015, February 10). WeChat's impact: A report on WeChat platform data. Retrieved August 15, 2016, from technode, http://technode.com/2015/02/10/wechat-impact-report/

Cowie, J., Oteniya, L. and Coles, R., 2007. Particle swarm optimisation for learning Bayesian networks. In *ICCIIS 2007, World Congress on Engineering, WCE 2007* (pp. 71-76). Newswood Limited/International Association of Engineers (IAENG).

De Campos, L.M. and Huete, J.F., 2000. A new approach for learning belief networks using independence criteria. *International Journal of Approximate Reasoning*, *24*(1), pp.11-37.

Elwert, F., 2013. Graphical causal models. In *Handbook of causal analysis for social research* (pp. 245-273). Springer Netherlands.

Falaki, H., Lymberopoulos, D., Mahajan, R., Kandula, S. and Estrin, D., 2010, November. A first look at traffic on smartphones. *In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (pp. 281-287). ACM.

Garg, R. and Telang, R., 2012. Inferring app demand from publicly available data. *MIS Quarterly*, *Forthcoming*.

Ghose, A. and Han, S.P., 2014. Estimating demand for mobile applications in the new economy. *Management Science*, *60*(6), pp.1470-1488.

Heckerman, D., Geiger, D. and Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, *20*(3), pp.197-243.

Huang, Y. and Valtorta, M., 2006, July. Identifiability in causal Bayesian networks: A sound and complete algorithm. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE* (Vol. 21, No. 2, p. 1149). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Kabli, R., Herrmann, F. and McCall, J., 2007, July. A chain-model genetic algorithm for Bayesian network structure learning. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation* (pp. 1264-1271). ACM.

Kalisch, M. and Bühlmann, P., 2007. Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*,*8*(Mar), pp.613-636.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M.H. and Bühlmann, P., 2012. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, *47*(11), pp.1-26.

Koller, D. and Friedman, N., 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

Larrañaga, P., Murga, R., Poza, M. and Kuijpers, C., 1996. Structure learning of Bayesian networks by hybrid genetic algorithms. In *Learning from Data* (pp. 165-174). Springer New York.

Maathuis, M.H. and Colombo, D., 2015. A generalized back-door criterion. *The Annals of Statistics*, *43*(3), pp.1060-1088.

Neapolitan, R.E., 2004. Learning Bayesian networks.

Pearl, J., 1982, August. Reverend Bayes on inference engines: A distributed hierarchical approach. In *AAAI* (pp. 133-136).

Pearl, J., 1993. [Bayesian Analysis in Expert Systems]: Comment: Graphical Models, Causality and Intervention. *Statistical Science*, *8*(3), pp.266-269.

Pearl, J., 1995. Causal diagrams for empirical research. *Biometrika*, *82*(4), pp.669-688.

Pearl, J., 2011. Bayesian networks. Department of Statistics, UCLA,

Peng, J., Wang, P., Zhou, N. and Zhu, J., 2012. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*.

Perković, E., Textor, J., Kalisch, M. and Maathuis, M.H., 2015. A complete generalized adjustment criterion. *arXiv preprint arXiv:1507.01524*.

Ramsey, J., Zhang, J. and Spirtes, P.L., 2012. Adjacency-faithfulness and conservative causal inference. *arXiv preprint arXiv:1206.6843*.

Richardson, T. and Spirtes, P., 2002. Ancestral graph Markov models. *Annals of Statistics*, pp.962-1030.

Sobel, M.E., 1996. An introduction to causal inference. *Sociological Methods & Research*, *24*(3), pp.353-379.

Schwarz, G., 1978. Estimating the dimension of a model. *The annals of statistics*, *6*(2), pp.461-464.

Scutari, M. and Denis, J.B., 2014. *Bayesian networks: with examples in R*. CRC Press.

Shpitser, I. and Pearl, J., 2008. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, *9*(Sep), pp.1941-1979.

Spirtes, Peter, Clark N. Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.

Tian, J. and Pearl, J., 2002, August. A general identification condition for causal effects. In *AAAI/IAAI* (pp. 567-573).

Tongaonkar, A., Dai, S., Nucci, A. and Song, D., 2013, March. Understanding mobile app usage patterns using in-app advertisements. *In International Conference on Passive and Active Network Measurement* (pp. 63-72). Springer Berlin Heidelberg.

Tsamardinos, I., Brown, L.E. and Aliferis, C.F., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine learning*,*65*(1), pp.31-78.

Pearl, T.V.J., 1991. Equivalence and synthesis of causal models. In *Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence* (pp. 220-227).

Wang, T., Touchman, J.W. and Xue, G., 2004, August. Applying two-level simulated annealing on Bayesian structure learning to infer genetic networks. In *Computational Systems Bioinformatics Conference*, *2004. CSB 2004. Proceedings. 2004 IEEE* (pp. 647-648). IEEE.

Xu, Y., Lin, M., Lu, H., Cardone, G., Lane, N., Chen, Z., Campbell, A. and Choudhury, T., 2013, September. Preference, context and communities: a multi-faceted approach to predicting smartphone app usage patterns. *In Proceedings of the 2013 International Symposium on Wearable Computers* (pp. 69-76). ACM.

Zhang, J., 2008. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, *9*(Jul), pp.1437-1474.

Zhang, J., 2008. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, *172*(16), pp.1873-1896.