Title
Data-Driven Optimization to Learn Structural Models

Author
Kaynar Keles, Sema Nur

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Data-Driven Optimization to Learn Structural Models

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Management

by

Sema Nur Kaynar Keles

2022

ABSTRACT OF THE DISSERTATION

Data-Driven Optimization to Learn Structural Models

by

Sema Nur Kaynar Keles

Doctor of Philosophy in Management

University of California, Los Angeles, 2022

Professor Auyon Adnan Siddiq, Chair

The rapid accumulation of high-dimensional data has opened new opportunities to make informed decisions. In this thesis, we focus on estimation of structural models from observational data using optimization and statistics to understand the effects of strategic decisions. We develop efficient procedures that blend techniques from economic modeling and machine learning to uncover underlying models efficiently and accurately.

In Chapter 2, we focus on understanding the effect of performance-based incentives on worker performance using historical contract data. The design of performance-based incentives can be naturally posed as a moral hazard principal-agent problem. In this setting, a key input to the principal's optimal contracting problem is the agent's production function – the dependence of agent output on effort. While agent production is classically assumed to be known to the principal, this is unlikely to be the case in practice. Motivated by the design of performance-based incentives, we present a method for estimating a principal-agent model from data on incentive contracts and associated outcomes, with a focus on estimating agent production. The proposed estimator is statistically consistent and can be expressed as a mathematical program. To circumvent computational challenges with solving the estimation problem exactly, we approximate it as an integer

program, which we solve through a column generation algorithm that uses hypothesis tests to select variables. We show that our approximation scheme and solution technique both preserve the estimator's consistency and combine to dramatically reduce the computational time required to obtain sound estimates. To demonstrate our method, we conducted an experiment on a crowdwork platform (Amazon Mechanical Turk) by randomly assigning incentive contracts with varying pay rates among a pool of workers completing the same task. We present numerical results illustrating how our estimator combined with experimentation can shed light on the efficacy of performance-based incentives.

In Chapter 3, we focus on learning causal structures from observational data, a process known as *causal discovery*. We propose a new optimization-based method for causal discovery. Our method takes as input observational data over a set of variables and returns a graph in which causal relations are specified by directed edges. We consider a highly general search space that accommodates latent confounders and feedback cycles, which few extant methods do. We formulate the discovery problem as an integer program, and propose a solution technique that leverages the conditional independence structure in the data to identify promising edges for inclusion in the output graph. Our method is among the very first to bring integer programming to general causal discovery, which we believe is one of our main contributions. In the large-sample limit, our method recovers a graph that is equivalent to the true data-generating graph. Computationally, our method is competitive with the state-of-the-art, and can solve in minutes instances that are intractable for alternative causal discovery methods. We then extend our framework to a priori identify a subset of variables that collectively carry all useful information about the variable of interest. This way, we can sidestep the computational burden of learning causal relations among variables of secondary importance.

In Chapter 4, we focus on investigating the validity of instrumental variables, which are widely used to estimate causal effects in the presence of unmeasured confounding. In particular, we

apply our method developed in Chapter 3 to US Census data from the seminal paper on the returns to education by (Angrist and Krueger, 1991), which contains a pioneering application of an instrumental variable, but one whose validity has been contested. We find that the causal structures uncovered by our method are consistent with the literature on the instrument from (Angrist and Krueger, 1991), and that our method pinpoints some of the sources of debate. Our results suggest that our graphical approach can be a useful complement to well-established empirical methods.

The dissertation of Sema Nur Kaynar Keles is approved.

Charles J. Corbett

Frederick Eberhardt

Elisa F. Long

Velibor Mišić

Auyon Adnan Siddiq, Committee Chair

University of California, Los Angeles

2022

This dissertation is dedicated to

the loving memory of my grandfather Mehmet Alıç whom I miss every day.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I am grateful to my advisor Auyon Siddiq for his continuous support over the last four years. Auyon was always there to guide me along the way. His positivity and unshakable belief in me made my years at UCLA very pleasant. He is the one who taught me the importance of working on questions that I am passionate about and being creative as a researcher. I would not be where I am today if it were not for his constant support and guidance.

I have been extraordinarily lucky to have Frederick Eberhardt as my mentor for the last three years of my PhD. He has been extremely supportive and extremely critical at the same time, always pushing me to ask the most important questions and helping me immensely while figuring them out. I would also like to thank Elisa Long for all the invaluable lessons that I learned from her over the years. Her passion on conducting research with real world impact has always inspired me. I also thank the other members of my dissertation advisory committee – Velibor Mišić and Charles Corbett – for providing critical feedback on this research and for their support on the academic job market.

I have also been fortunate to have a wonderful cohort of researchers at UCLA. I am grateful for Yi-Chun Chen for his friendship and support that I will miss very much. I would like to thank my dear friends who made DOTM office incredibly warm and the atmosphere very supportive. In particular, thank you Pankaj Jindal, İrem Akçakuş, Mirel Yavuz, Jingwei Zhang, Saeed Ghodsi, Jingyuan Hu, Zach Siegel, Xinyi Guan, Jian Gao, Martín González Cabello, and Abolfazl Taghavi.

I would like to thank my family for their immense love and support. I am thankful for my parents Nuray and Mehmet for loving and taking care of me, for providing a peaceful home to grow in, and for making sure I get a great education. I would like to thank my sister Kutsal for being my life long partner in crime and for always being close in heart even when we are thousands of miles apart. I would also like to thank Nilay Korgal for being my best friend and for supporting

me through thick and thin.

I would like to thank my husband Umit Keles for being the most loving, caring, thoughtful, and intelligent person that I have ever met. None of this would be possible without you. There are no words to express how grateful I am to have you in my life. Our life together is the greatest thing I could have asked for. Last but not least, I am also thankful for Venice. Seeing her paws and little face warms my heart so deeply.

Vita

EDUCATION

University of California, Los Angeles, 2022
Ph.D. Candidate in Management (Decisions, Operations & Technology Management)

Bilkent University, 2017
B.Sc., Industrial Engineering

Bilkent University, 2015
M.Sc., Industrial Engineering

PUBLISHED & UNDER REVIEW PAPERS

Frederick Eberhardt, **Nur Kaynar**, Auyon Siddiq. "Discovering Causal Models with Optimization: Confounders, Cycles, and Feature Selection." Major revision, *Management Science*.

Nur Kaynar, Auyon Siddiq. "Estimation of a Non-Parametric Principal-Agent Model with Hidden Actions." *Management Science* (2022). Forthcoming.

Kelsea Best, Michelle Miro, Rachel Kirpes, **Nur Kaynar**, and Aisha Najera Chesler. "Data-driven decision support tools for assessing the vulnerability of community water systems to groundwater contamination in Los Angeles County." *Environmental Science & Policy.* 393-400, 2021.

Michelle Miro,Kelsea Best, **Nur Kaynar**, Rachel Kirpes, and Aisha Najera Chesler. "Approaches to analyzing the vulnerability of community water systems to groundwater contamination in Los Angeles County." *Research in Mathematics and Public Policy.* Springer, Cham, 19-28, 2020.

Nur Kaynar, Özlem Karsu. "Equitable decision making approaches over allocations of multiple benefits to multiple entities." *Omega.* 81:85-98, 2017.

Chapter 1

Introduction

Understanding effects of strategic decisions is at the core of operations research and management science. In this thesis, we develop new methodologies that combine optimization with economic modeling and statistics to uncover the mechanisms that govern the effects of decisions. The chapters are unified around the central motive of learning these mechanisms, but the methodologies used within them are diverse. In Chapter 2, we use an agent-based modeling to understand the effects of performance-based incentives on worker production. Whereas in Chapter 3, we built on a recently emerged framework that provides a theoretical foundation to uncover underlying causal relations from probability distributions without needing to specify a model. In Chapter 4, we discuss how these newly emerged automated ways of inferring causation from data can be used to gain more insights for validity of instrumental variables, which are widely used to estimate causal effects in the presence of unmeasured confounding.

Next, we provide a high-level overview of the main ideas contained within each chapter, including their main contributions to the literature.

1.1 Estimating Effects of Incentive Contracts in Online Labor Platforms

Chapter 2 of this thesis focuses on analysis of incentive contracts in online labor platforms. Principal-agent models are widely used to study contracting problems. In a typical setting, the principal delegates work or decision-making authority to an agent that acts in their own best interest, and whose incentives are misaligned with the principal's. Further, there usually exists some form of information asymmetry between the two parties, where the principal cannot observe either the agent's actions or other private information. The agent is compensated by the principal based on an observable outcome, which depends on the action taken by the agent. The principal has preferences over the outcomes, but wields little direct influence over them. As a consequence, the contracting problem faced by the principal is to design a compensation scheme for the agent that induces desirable outcomes.

Despite their prominence in the operations literature, the *estimation* of principal-agent models – that is, extracting an agent's decision-making model from observational data – has received little attention to date. Our key contribution is to propose an intuitive estimator for a general class of principal-agent models where agent actions are hidden. Our agent model is non-parametric, and imposes minimal structural assumptions on the agent's utility function. We present a mild condition under which the model is identifiable, and show the estimator to be statistically *consistent*. We show that the estimation problem with hidden actions is $\mathcal{NP}$-hard, contrary to the case where actions are observable.

We then show that, with a slight modification, the estimator can be formulated exactly as a mixed-integer linear program. However, solving the exact estimator using off-the-shelf optimization solvers is non-viable for larger instances. To overcome this deficiency, we propose an approximation

scheme based on minimizing the loss function over a discrete subset of the original parameter space. To solve the restricted estimator, we propose an iterative solution algorithm in the spirit of column generation that exploits statistical properties of the integer programming formulation. The defining feature of our algorithm is the use of two-sample, non-parametric hypothesis tests to select variables to introduce into the model. Critically, we show that the proposed solution algorithm preserves the statistical consistency of the exact estimator. We further show that the expected iteration count of the algorithm can be bounded by a function of the chosen significance level used in the hypothesis testing step. To the best of our knowledge, this algorithm is the first to incorporate hypothesis tests into a solution technique for integer programs.

We showcase the efficacy of the estimator in two sets of numerical experiments. First, we use synthetic instances to show that the statistical column generation algorithm produces solutions that are competitive with the exact estimator in a fraction of the computational time. To demonstrate our method, we conducted an experiment on a crowd-work platform (Amazon Mechanical Turk) by randomly assigning incentive contracts with varying pay rates among a pool of workers completing the same task. We present numerical results illustrating how our estimator combined with experimentation can shed light on the efficacy of performance-based incentives in such online labor platforms.

1.2 Discovering Causal Models with Optimization

In Chapter 3, we consider how to learn causal relations from observational data. This inference is known as *causal (structure) discovery*. While randomized experiments are often considered the gold standard for identifying causal relations, they come with substantial limitations: The experimenter has to be able to fully control the treatment, or adjust for the failure to do so. This often implies that the experiments have to be conducted in artificial environments with small sample sizes,

undermining their validity. Further, some interventions are very costly to perform, and some are unethical. Consequently, it is often desirable to learn as much as possible about underlying causal structures from observational data alone, without performing experiments.

Our main contribution is a new optimization-based method for causal discovery that allows for both unmeasured confounders *and* feedback cycles. Our method takes as input observational data over a set of variables, and returns a graph in which causal relations are specified by directed edges. There are very few extant discovery methods that consider this extremely general search space, and those that do, do not scale well. To achieve better scalability, we propose a solution approach that exploits the conditional (in)dependence structure in the data to detect "promising" candidate edges in the underlying graph, which are then assembled into a causal graph by an optimization model. We computationally show that our approach allows us to solve in minutes instances that are outright intractable for recently proposed methods. Our main theoretical result is to show that this technique asymptotically recovers a graph that is equivalent to the true, data-generating graph.

The proposed methodology constructs a complete causal graph over a set of observed variables, in which all causal relations are deemed equally important. In empirical applications, however, there may be a target variable whose causes are of particular interest to the researcher (e.g., income or health outcomes). We then extend our framework to a priori identify a subset of variables that collectively carry all useful information about the variable of interest using optimization. This way, we can sidestep the computational burden of learning causal relations among variables of secondary importance.

1.3 Graphical Validation of Instrumental Variables

In Chapter 4, we focus on investigating the validity of an instrumental variable using the method developed in Chapter 3. Instrumental variables are widely used to estimate causal effects primarily

to overcome bias in regression estimates of causal effects that stem from unmeasured confounding. The use of instrumental variables comes with its own challenges, namely, identifying an appropriate instrument in the first place. Even though the conditions for a variable to be a valid instrument are well established (e.g., Angrist et al. (1996)), testing its validity from data remains a challenge as these conditions are largely believed to be untestable from data. For the most part, instruments are selected subjectively based on domain knowledge and verbal justifications.

Previous work (Pearl, 2000) discusses that these conditions can be expressed in a graphical framework in terms of presence or absence of particular edges and paths. Building on this correspondence, we demonstrate how the method proposed in Chapter 3 can be used to shed light on the validity of a proposed instrument in an objective, automated, and intuitive manner. Specifically, within our graphical framework, checking instrument validity amounts to learning causal structures over the relevant data and checking the extent to which the output graph abides by graphical criteria on instrumental validity.

We apply our method to the well-known dataset on educational attainment and income from Angrist and Krueger (1991), which contains one of the most influential applications of an instrumental variable, but one whose validity has been the subject of debate. In that work, the authors propose quarter-of-birth as an instrument for educational attainment. To the best of our knowledge, our work is the first to examine the quarter-of-birth instrument from a causal discovery perspective. Although the graphical framework we use here is quite different from previous investigations of the instrument from Angrist and Krueger (1991), our main finding is remarkably consistent with the literature – that the quarter-of-birth instrument is plausible, but weak.

Chapter 2

Estimating Effects of Incentive Contracts in Online Labor Platforms

2.1 Introduction

The extent to which financial incentives increase worker performance is of interest in many employment settings. This question has taken on renewed relevance due to the emergence of online labor platforms, which are used for on-demand jobs like ride-hailing (e.g., Uber, Lyft), delivery (Postmates), freelance work (Upwork), and short, discrete tasks (Amazon Mechanical Turk). Although these platforms support different types of work, they also have common features: workers are hired and compensated on a per-task basis, work is done remotely with limited supervision, and workers may be offered performance-based incentives.[1]

The design of performance-based incentives can be naturally posed as a moral-hazard principal-agent problem, in which an agent's (worker's) effort is hidden from the principal (employer), and the agent's output depends stochastically on their effort (Holmstrom, 1979; Grossman and Hart, 1983;

[1]For example, Lyft offers drivers bonuses for fulfilling a target number of rides within a predefined time frame (Lyft, 2021), and Postmates offers a similar incentive (Postmates, 2021). Similarly, freelance platforms Upwork and Amazon Mechanical Turk allow clients to provide workers with bonuses at their own discretion.

Sappington, 1991). In this setting, the relationship between worker output and effort corresponds to a set of parameters that define agent production. If these parameters are known, then the principal's problem of optimally designing incentives is well-defined and potentially convex (Grossman and Hart, 1983).

In practice, however, the relationship between worker effort and output is unlikely to be known *a priori*. Given data on incentives and associated output, this dependence can be inferred by specifying an appropriate agent model and estimating the parameters that govern agent production. Despite the importance of principal-agent models to the analysis of incentive contracts, estimation problems of this nature are scarce in the literature, even for simple agent models. Estimating an agent model from observational or experimental data can be a useful step toward the design of incentive contracts in practice, and can also play a role in estimating agent welfare under a given contract.

Our main contribution is to present an estimator for a principal-agent model with hidden actions, along with an algorithm for solving the estimation problem. Our focus is on estimating model parameters that encode agent production, namely, the conditional distribution over output for each effort level. To reflect a moral-hazard setting, we assume no data is available on agent effort, which makes the estimation problem computationally non-trivial. We make two methodological contributions in particular: (1) we provide an estimator that is statistically *consistent* under appropriate conditions, meaning it uncovers the true model parameters as the sample size goes to infinity, and (2) we develop an accompanying solution technique that is computationally efficient and preserves consistency.

The agent model we consider is non-parametric, in that we do not assume functional forms for the dependence of agent output on effort, and we assume both output and effort levels are discrete. This specification has two important consequences. First, it admits a simple and *tractable*

formulation of a general optimal contracting problem, which allows us to readily solve for an optimal contract under the estimated agent model. Second, estimating agent models is well-known to be challenging due to a need to embed the agent's problem – itself an optimization problem – within the estimator (Bajari et al., 2007). Our modeling approach allows us to express the estimator as an integer program, which admits a structure that supports obtaining estimates quickly using a novel solution technique. In addition to these computational advantages, our non-parametric model naturally handles threshold-based incentives, which commonly arise in practice, and is flexible enough to have strong predictive performance on a variety of datasets without overfitting.

In an empirical study, we show how our estimator can be combined with experimentation to characterize worker output over a class of incentive contracts, which in turn allows us to solve for an optimal contract from the given class. In a randomized experiment, we recruited a pool of 500 workers from a crowdwork platform (Amazon Mechanical Turk), each of whom was asked to complete an identical proofreading task, with output measured by the number of typos identified. We created exogenous variation in payments by randomly generating the parameters of an incentive contract for each worker. We then applied our estimator to the experimental data to investigate the effect of performance-based incentives on worker output. Our results complement existing findings that incentives do increase output in crowdwork, although we observe diminishing returns to output beginning at relatively low payments.

Our model has limitations. The agent model does not include common features of principal-agent problems; in particular, we do not address risk aversion or unobserved agent heterogeneity in this work. This abstraction arises from our focus on obtaining consistent estimates (potentially for a large number of parameters) while maintaining computational tractability. Generalizing our estimation procedure to accommodate a richer class of agent models may expand its applicability in practice. Further, our non-parametric approach may be unsuitable for settings with limited data,

because it may require estimating many parameters if the action or outcome space is large.

The remainder of this chapter is organized as follows. §2.2 defines the agent model, presents the associated estimator, and establishes consistency. §2.3 presents an exact formulation of the estimator as an integer program and discusses the computational challenges of the exact representation. §2.4 develops an approximate estimator and an accompanying solution technique, which dramatically improve tractability while preserving consistency of the exact estimator. §2.5 describes the randomized experiment and demonstrates the application of our estimator to experimental data. §2.6 concludes. All proofs are contained in Appendix *A*.

2.1.1 Related literature

Existing work on estimating principal-agent models is relatively limited. Several papers have focused on employee compensation. Ferrall and Shearer (1999) use payroll records of copper mine workers to estimate the cost of employee risk aversion. Paarsch and Shearer (2000) use a tree-planting firm's records to estimate the impact of providing piece-rate compensation over fixed wages, and Shearer (2004) addresses the same question through a field experiment. Duflo et al. (2012) estimate an agent model to assess the impact of financial incentives for schoolteachers and use the model to estimate cost reductions associated with a counterfactual payment scheme. Misra et al. (2005) and Misra and Nair (2011) both estimate agent models based on salesforce compensation and empirically validate the models on out-of-sample data. Gayle and Miller (2015) focus on identifying a general principal-agent model motivated by managerial compensation. Georgiadis and Powell (2021) provide conditions under which a single A/B test can estimate the impact of marginal changes to an incentive contract, using the classical principal-agent model from Holmstrom (1979). Applications beyond employee compensation include agriculture (de Zegher et al., 2017) and healthcare (Vera-Hernandez, 2003; Lee and Zenios, 2012; Aswani et al., 2019).

Previous work on estimating principal-agent models have used a variety of methods, including least squares (Lee and Zenios, 2012), simulated method of moments (Paarsch and Shearer, 2000; Misra et al., 2005; Misra and Nair, 2011; Duflo et al., 2012), simulation-based maximum likelihood estimation (Ferrall and Shearer, 1999; Vera-Hernandez, 2003; Aswani et al., 2019), and numerical minimization of a sum-of-squares criterion (Gayle and Miller, 2015). Our approach differs in that we formulate the estimation problem as an integer program, which is made possible by our specification of the agent model, in particular by assuming agent actions and outputs are discrete.

We solve the estimation problem using a column generation algorithm that exploits statistical properties of the formulation. Column generation methods have been successfully applied to solve large-scale linear and integer programs in which an extremely large number of variables is the main obstacle to obtaining optimal solutions (Vanderbeck and Wolsey, 1996; Barnhart et al., 1998; Lubbecke and Desrosiers, 2005). These methods typically involve solving a tractable *master* problem that restricts attention to a subset of decision variables, and selectively introducing variables into the formulation until a certificate of optimality or alternative termination criterion is met. In contrast to existing column generation methods that select columns using dual information, our algorithm uses a series of non-parametric hypothesis tests to identify variables to introduce into the master problem. This approach is viable in our setting because the decision variables are mapped to empirical probability distributions constructed from the data, giving them a clear statistical interpretation. By comparison, existing column generation methods have typically been applied to deterministic settings where the model parameters may not have any statistical meaning (see Lubbecke and Desrosiers (2005) for a review).

The estimation problem we consider is also closely related to a recent line of research on *inverse* optimization, in which optimization model parameters are inferred from (potentially noisy) solution data. Existing approaches to inverse optimization have focused on estimating parameters

of linear programs (Chan et al., 2018) or general convex optimization problems (Keshavarz et al., 2011; Bertsimas et al., 2015; Aswani et al., 2018). Similar to our work, the literature on inverse optimization is often motivated by an interest in estimating a model of agent decision-making from data (Aswani et al., 2018; Esfahani et al., 2018). Our approach differs in that instead of assuming the agent solves a convex optimization problem, we assume they select a utility-maximizing action from a discrete set, which calls for a different solution approach.

2.1.2 Notation

For convenience, we describe notational conventions here. Sets are denoted by upper case letters, scalars by lower case letters, and vectors and matrices by lower case, boldfaced letters. For a $m \times d$ matrix $\mathbf{x}$, let $\mathbf{x}_a$ be the a^{th} row, and let x_{aj} be the entry in the a^{th} row and j^{th} column. For vectors $\mathbf{x}$ and $\mathbf{y}$, let $\|\mathbf{x}\|_1 = \sum_{a=1}^{m} \sum_{j=1}^{d} |x_{aj}|$ denote the ℓ_1-norm, and let $\mathbf{x} \circ \mathbf{y} = \sum_{a=1}^{m} \sum_{j=1}^{d} x_{aj} y_{aj}$ be the elementwise product. For a matrix of random variables $\mathbf{x}_n$, we use both $\mathbf{x}_n \longrightarrow \mathbf{x}^0$ and $\text{plim}_{n\to\infty} \mathbf{x}_n = \mathbf{x}^0$ to mean $\mathbf{x}_n$ converges elementwise in probability to $\mathbf{x}^0$ as $n \longrightarrow \infty$, unless otherwise specified. Define the indicator variable $\mathbb{I}\{\cdot\} = 1$ if the statement $\{\cdot\}$ is true, and 0 otherwise. For simplicity, we use $\mathbb{E}(\cdot)$ for all expectations and $\Pr(\cdot)$ for all probabilities throughout the chapter.

2.2 Estimator

In this section, we define the principal-agent model (§2.2.1), formulate the estimator (§2.2.2), and prove its statistically consistency (§2.2.3).

2.2.1 Principal-agent model and contract data

Our principal-agent model is a discrete analogue to the canonical model introduced by Grossman and Hart (1983). We choose this model for both its simplicity and generality. The interaction between the principal and agent proceeds as follows. The principal selects a *contract* to offer the agent, which is a mapping of payments to *outcomes* (i.e., agent output). Outcomes depend stochastically on a costly *action* (i.e., effort) taken by the agent. Outcomes are observed by both parties, while actions are observed only by the agent.

Let A and J index the set of possible actions and outcomes, respectively, where $|A| = m$ and $|J| = d$. Let ξ be a discrete random variable denoting the outcome, where $\xi \in J$. We denote a contract by $\mathbf{r} \in \mathbb{R}_+^d$, where r_j is the payment to the agent if outcome j is realized. Let $\mathbf{c} \in \mathbb{R}_+^m$ denote action costs, where c_a is the cost to the agent of taking action a. The dependence of outcomes on actions is governed by a parameter matrix $\boldsymbol{\pi} \in \mathbb{R}_+^{m \times d}$, where π_{aj} denotes the probability that action a leads to outcome j. We use $\boldsymbol{\pi}_a \in \mathbb{R}^d$ to denote the probability mass function over outcomes associated with action a.

Given a contract $\mathbf{r}$, the agent selects an action to maximize their expected utility by solving

$$\max_{a \in A} \left\{ \sum_{j \in J} \pi_{aj} r_j - c_a \right\}. \tag{2.1}$$

We assume that there exists at least one action that yields non-negative expected utility for the agent. If for each $a \in A$ the distribution $\boldsymbol{\pi}_a$ is known, the principal's problem of selecting a utility-maximizing contract can be formulated as a convex optimization problem (Grossman and Hart, 1983). We take an inverted perspective in this work, by instead supposing that the distributions $\boldsymbol{\pi}_a$, $a \in A$ are unknown, but may be estimated given appropriate data. In particular, suppose we

have data from n identical agents,[2]

$$(\mathbf{r}^i, \xi^i),\ i \in I, \tag{2.2}$$

where I indexes pairs of incentive contracts and outcomes, and $|I| = n$. Let $R \subseteq \mathbb{R}^d$ be the set of all possible values of $\mathbf{r}^i$. Further, we assume the contract set R is bounded, in that there exists a constant $\bar{r}$ such that $\bar{r} = \sup_{\mathbf{r}\in R} \|\mathbf{r}\|_0 < \infty$. The assumption that R is bounded ensures that the contracts $\mathbf{r}^i$ remain bounded as $n \longrightarrow \infty$.

Next, suppose we have no observations of past agent actions, and only know the agent's action set A and associated costs, $\mathbf{c}$. A natural question in this setting is to predict the distribution of the outcome ξ^{n+1} under a new contract $\mathbf{r}^{n+1}$. Note that if $\boldsymbol{\pi}$ is known, then this prediction task reduces to solving the agent's problem (2.1) under $\mathbf{r}^{n+1}$, identifying the optimal action a, and taking $\boldsymbol{\pi}_a$ to be the distribution of ξ^{n+1}. Therefore, the matrix $\boldsymbol{\pi}$ is the key model primitive for predicting the outcome associated with $\mathbf{r}^{n+1}$. Our goal is to estimate the parameter $\boldsymbol{\pi}$ from data that takes the form given in (2.2).

The assumption that agent costs are unknown is relatively mild in our setting, given that agent actions are also hidden. From a model-fitting perspective, it suffices to select $\mathbf{c}$ to cover a range of *possible* costs to the agent. In our numerical study in §2.5, we take a machine learning perspective by treating the number of agent actions m and the set of costs $\mathbf{c}$ as hyperparameters that are tuned prior to fitting the model.

2.2.2 Estimator formulation

Next, we formalize the estimator for $\boldsymbol{\pi}$.[3] Let

[2] We extend our model to accommodate heterogeneous agents in §A.3 of Appendix.

[3] Throughout this chapter, we shall use *estimator* to refer to an optimization problem or algorithm, and *estimate* to refer to its solutions.

$$A(\mathbf{r}, \boldsymbol{\pi}) = \underset{a \in A}{\operatorname{argmax}} \left\{ \sum_{j \in J} \pi_{aj} r_j - c_a \right\} \tag{2.3}$$

denote the set of optimal actions under the contract $\mathbf{r}$ and the model $\boldsymbol{\pi}$. Let $\mathbf{y} \in \{0,1\}^{n \times d}$ be a binary matrix that encodes historical outcomes, where $y_j^i = 1$ if $\xi^i = j$ and $y_j^i = 0$ if $\xi^i \neq j$. For each $i \in I$, let x^i be a decision variable representing the agent action under contract $\mathbf{r}^i$, and let $\boldsymbol{\omega} \in \mathbb{R}_+^{m \times d}$ be a set of auxiliary variables, which will be used to model empirical probabilities. For fixed $\boldsymbol{\pi}$, the loss function $L_n(\boldsymbol{\pi})$ is then given by

$$L_n(\boldsymbol{\pi}) = \underset{\mathbf{x}, \boldsymbol{\omega}}{\text{minimize}} \ \sum_{a \in A} \sum_{j \in J} |\pi_{aj} - \omega_{aj}| \tag{2.4a}$$

$$\text{subject to } \ x^i \in A(\mathbf{r}^i, \boldsymbol{\pi}), \quad i \in I, \tag{2.4b}$$

$$\omega_{aj} = \frac{1}{|\{i | x^i = a\}|} \sum_{i \in \{i | x^i = a\}} y_j^i, \quad a \in A, \ j \in J. \tag{2.4c}$$

In the formulation above, (2.4b) restricts each x^i to be an optimal action under $\mathbf{r}^i$ and $\boldsymbol{\pi}$, and (2.4c) defines ω_{aj} to be the empirical probability that action a leads to outcome j. Note that the empirical probability ω_{aj} depends on the cardinality of the set $\{i | x^i = a\}$, which is the implied number of data points for which the action a is optimal for the agent under $\boldsymbol{\pi}$. The objective (2.4a) then simply measures the error between the model probabilities $\boldsymbol{\pi}$ and the implied empirical probabilities $\boldsymbol{\omega}$.

Next, let Π be a compact set representing the parameter set for $\boldsymbol{\pi}$. The estimate is then attained at a minimizer of the loss function over Π:

$$\text{(PA)} \qquad \hat{\boldsymbol{\pi}}_n \in \underset{\boldsymbol{\pi} \in \Pi}{\operatorname{argmin}} \ L_n(\boldsymbol{\pi}).$$

It will be convenient to interpret the parameter set Π as the Cartesian product of m probability

simplices – one for each action $a \in A$.

2.2.3 Statistical consistency

Let us now suppose there exists a "true" model parameter $\boldsymbol{\pi}^0$ that is responsible for generating the data $(\mathbf{r}^i, \xi^i)$, $i \in I$. We say an estimator is *statistically consistent* if it produces a sequence of estimates $\hat{\boldsymbol{\pi}}_n$ such that $\hat{\boldsymbol{\pi}}_n \longrightarrow \boldsymbol{\pi}^0$ as $n \longrightarrow \infty$. This raises a natural question: Under what conditions, if any, is PA a consistent estimator? In general, whether an estimator is consistent depends on the specification of the loss function. Our main result in this section, Theorem 1, shows that minimizing the loss function $L_n(\boldsymbol{\pi})$ defined in (4) produces an estimate that is indeed consistent.

Before addressing the consistency of PA, we first formalize the statistical model that generates the data. First, we define an important set that is used throughout our analysis:

$$R_a(\boldsymbol{\pi}) = \left\{ \mathbf{r} \in R \middle| a \in \operatorname*{argmax}_{a \in A} \sum_{j \in J} \pi_{aj} r_j - c_a \right\}, \tag{2.5}$$

Here, $R_a(\boldsymbol{\pi})$ represents the subset of the contract set R where action $a \in A$ is optimal for the agent, given the model $\boldsymbol{\pi}$. Next, we impose two assumptions. The first assumption formalizes the data generation process.

Assumption 1 (Data). *The data $(\mathbf{r}^i, \xi^i)$, $i \in I$, are independent samples of random variables $(\mathbf{r}, \xi)$, where (i) $(\mathbf{r}, \xi)$ are jointly distributed with support $R \times J$, (ii) $\mathbf{r}$ has continuous marginal density function $f(\mathbf{r})$, (iii) $\Pr(\mathbf{r} \in R_a(\boldsymbol{\pi})) > 0$ for all $a \in A$ and $\boldsymbol{\pi} \in \Pi$, and (iv) ξ has conditional mass function $\pi^0_{aj} = \Pr(\xi = j | \mathbf{r} \in R_a(\boldsymbol{\pi}^0))$, where $\boldsymbol{\pi}^0 \in \Pi$.*

Assumption 1(iv) states that there exists a "true" parameter – denoted $\boldsymbol{\pi}^0$ – that is responsible for generating the outcomes ξ^i, based on the agent model (2.1). The statements in (ii) and (iii)

are regularity conditions that we use to prove convergence of $\hat{\boldsymbol{\pi}}_n$ to $\boldsymbol{\pi}^0$.[4] Our assumption that the data are independent and identically distributed (i.i.d.) is commonly used in the statistical learning literature to obtain similar consistency results.[5]

Next, we consider an additional condition that is important for our main result in Theorem 1.

Assumption 2 (Identifiability). *For every $\boldsymbol{\pi} \in \Pi$ such that $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$, there exists an (a, j) such that*

$$\pi_{aj} \neq \sum_{b \in A} \pi^0_{bj} \cdot \Pr(\mathbf{r} \in R_b(\boldsymbol{\pi}^0) | \mathbf{r} \in R_a(\boldsymbol{\pi})).$$

Assumption 2 is an *identifiability* condition, which ensures that the unknown parameter $\boldsymbol{\pi}^0$ can be learned from the data. This assumption implies a one-to-one mapping between the parameter set Π and the joint distribution of the random variables $(\mathbf{r}, \xi)$. In other words, Assumption 2 guarantees that the distribution of $(\mathbf{r}, \xi)$ is unique for each $\boldsymbol{\pi} \in \Pi$. In the absence of model identifiability, there may exist multiple parameters values in Π that generate the same distribution in the observed data; in this case, it is impossible for any estimation procedure to pinpoint the true $\boldsymbol{\pi}^0$. Identifiability conditions like Assumption 2 are commonly imposed to prove consistency of an estimator (Van der Vaart, 2000).

We can now present the main result of §2.2, which shows that the estimator PA uncovers the true model parameter $\boldsymbol{\pi}^0$ under Assumptions 1 and 2.

Theorem 1. *Let Assumption 1 hold. Then $\hat{\boldsymbol{\pi}}_n \longrightarrow \boldsymbol{\pi}^0$ for any $\boldsymbol{\pi}^0 \in \Pi$ if and only if Assumption*

[4] The assumption that the contract data $\mathbf{r}^i$, $i \in I$ is generated by a continuous density function $f(\mathbf{r})$ is important for our technical results. Intuitively, because the $\mathbf{r}^i$ are input data, assuming this continuity provides the estimator with more information, which makes precise inference of $\boldsymbol{\pi}^0$ possible under the identifiability condition in Assumption 2. If the contract data is instead generated by a discrete distribution supported on a subset of R, then a stronger identifiability than Assumption 2 is needed to compensate for the loss of information. We consider such a case in §A.4 of the electronic companion.

[5] Because we assume the data is generated by n independent agents making decisions simultaneously, which is plausible in online labor platforms, the i.i.d. assumption is not particularly restrictive for our setting. Moreover, this assumption is not strictly necessary to achieve consistency, depending on the problem setup. In §A.4 of the electronic companion, we consider a variation of the model where $\boldsymbol{\pi}^0$ can be estimated by dynamically selecting the contracts to offer the agent. This breaks the independence assumption on the contracts $\mathbf{r}^i$, but allows for consistent estimation of $\boldsymbol{\pi}^0$ under a different set of assumptions.

2 holds.

Theorem 1 states that the estimator PA is statistically *consistent*, which is defined as the convergence of estimates to the true model parameters (Van der Vaart, 2000; Casella and Berger, 2002; Bickel and Doksum, 2015). Despite being an asymptotic property, consistency is valuable in practice, because it guarantees that parameter estimates will generally improve with additional data. Conversely, an inconsistent estimator may produce inaccurate estimates of the unknown parameters, even if data is abundant. In pathological cases, the accuracy of an inconsistent estimator may even decrease with additional data. Therefore, a proof of consistency provides some assurance that parameter estimates will be "reasonable" under moderate sample sizes, and that the accuracy of the estimates will continue to improve with additional data.

Having established that the estimate $\hat{\boldsymbol{\pi}}_n$ behaves desirably, we now shift our attention to solving the estimator PA. Note that in a setting where agent actions are observable, a consistent estimate of $\boldsymbol{\pi}^0$ can be obtained by simply counting the relative frequency of outcomes associated with each action. In contrast, when agent actions are hidden, the estimation problem is non-trivial. At a high level, our approach for solving PA will be to leverage integer programming within a broader solution algorithm. The key challenge we face in solving PA is to develop a solution method that satisfies two criteria: (1) is computationally efficient, and (2) preserves the statistical consistency of PA. We note here that an alternative solution approach might be to formulate and solve a convex approximation to PA, although doing so is may result in an inconsistent estimator. We will therefore focus on obtaining solutions to PA directly.

2.3 Exact Integer Programming Formulation

In this section, we present an approach for solving PA exactly using integer programming. We will assume throughout that the parameter set Π is of the form

$$\Pi = \left\{ \boldsymbol{\pi} \in Q_\pi \middle| \boldsymbol{\pi} \geq 0, \sum_{j \in J} \pi_{aj} = 1 \text{ for } a \in A \right\}, \tag{2.6}$$

where Q_π is a polyhedron defined by a set of linear inequalities in $\boldsymbol{\pi}$. Assuming that $\boldsymbol{\pi} \in Q_\pi$ permits the formulation of the estimator as a mixed-integer linear program, while also allowing Π to capture various shape constraints on the parameter $\boldsymbol{\pi}$. For example, if

$$Q_\pi = \left\{ \boldsymbol{\pi} \middle| \sum_{k=j}^{d} \pi_{ak} \leq \sum_{k=j}^{d} \pi_{(a+1)k}, a \in \{1, 2, \ldots, m-1\}, j \in J \right\}, \tag{2.7}$$

then for any $a \in \{1, 2, \ldots, m-1\}$, Π forces the distribution $\boldsymbol{\pi}_{a+1}$ to stochastically dominate $\boldsymbol{\pi}_a$ in the first order, meaning costlier actions taken by the agent are more likely to generate high output. Alternatively, if $Q_\pi = \mathbb{R}^{m \times d}$, then Π permits each $\boldsymbol{\pi}_a$ to be any valid probability mass function over the outcomes J. We will assume throughout that Π satisfies (2.6) unless otherwise stated.

Although PA is based on an intuitive loss function, a naive formulation of PA as a mathematical program yields non-linear terms in the objective, due to how the variable $\boldsymbol{\omega}$ enters the loss expression (2.4a). However, the estimation problem is amenable to mathematical programming approaches under a slight modification. Consider the proxy loss function

$$\begin{aligned} Z_n(\boldsymbol{\pi}) = \underset{\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\omega}}{\text{minimize}} \;\; & \frac{1}{n} \sum_{a \in A} \sum_{j \in J} \eta_{aj} |\pi_{aj} - \omega_{aj}| \\ & \eta_{aj} = |\{i | x^i = a\}|, \quad a \in A, j \in J, \\ & \text{(2.4b)} - \text{(2.4c)}. \end{aligned} \tag{2.8}$$

Here, η_{aj} is the number of observations for which action a is implied to be optimal for the agent under $\boldsymbol{\pi}$. The loss function $Z_n(\boldsymbol{\pi})$ can be interpreted as a scaled version of $L_n(\boldsymbol{\pi})$, where the (a, j) component of $L_n(\boldsymbol{\pi})$ is scaled by η_{aj}/n. The proxy estimator is then given by

$$\boldsymbol{\pi}_n^* = \operatorname*{argmin}_{\boldsymbol{\pi} \in \Pi} Z_n(\boldsymbol{\pi}). \tag{2.9}$$

Next, we show that (2.9) can be formulated exactly as a mixed-integer linear program. With a slight abuse of notation, let $\mathbf{x} \in \{0,1\}^{m \times n}$ be binary variables, where $x_a^i = 1$ if $a \in A(\mathbf{r}^i, \boldsymbol{\pi})$, and $x_a^i = 0$ if $a \notin A(\mathbf{r}^i, \boldsymbol{\pi})$. Introducing the auxiliary variables z_{aj} to linearize the absolute values in the objective of (2.8) (Bertsimas and Tsitsiklis, 1997) yields the following formulation:

$$\begin{aligned}
\underset{\boldsymbol{\pi}, \mathbf{x}, \mathbf{z}}{\text{minimize}} \quad & \sum_{a \in A} \sum_{j \in J} z_{aj} && \text{(2.10a)} \\
\text{subject to} \quad & z_{aj} \geq \frac{1}{n} \sum_{i \in I} (y_j^i - \pi_{aj}) x_a^i \quad a \in A,\ j \in J, && \text{(2.10b)} \\
& z_{aj} \geq \frac{1}{n} \sum_{i \in I} (\pi_{aj} - y_j^i) x_a^i \quad a \in A,\ j \in J, && \text{(2.10c)} \\
& \left(\sum_{j \in J} \pi_{aj} r_j^i - c_a \right) x_a^i \geq \left(\sum_{j \in J} \pi_{bj} r_j^i - c_b \right) x_a^i, \quad i \in I,\ a \in A,\ b \in A, && \text{(2.10d)} \\
\text{(PA-C)} \quad & \sum_{j \in J} \pi_{aj} = 1, \quad a \in A, && \text{(2.10e)} \\
& \sum_{a \in A} x_a^i = 1, \quad i \in I, && \text{(2.10f)} \\
& x_a^i \in \{0,1\}, \quad a \in A, && \text{(2.10g)} \\
& \pi_{aj} \geq 0, \quad a \in A,\ j \in J, && \text{(2.10h)} \\
& \boldsymbol{\pi} \in Q_\pi. && \text{(2.10i)}
\end{aligned}$$

The objective (2.10a) and constraints (2.10b)–(2.10c) represent the error function $\frac{1}{n} \| \boldsymbol{\eta} \circ (\boldsymbol{\pi} - \boldsymbol{\omega}) \|_1$

given in (2.8). Constraint (2.10d) ensures that $x_a^i = 1$ only if $a \in A(\mathbf{r}^i, \boldsymbol{\pi})$, that is, only if action a is optimal under contract $\mathbf{r}^i$ and the parameter $\boldsymbol{\pi}$. Constraint (2.10e) ensures that the probability vector $\boldsymbol{\pi}_a$ sums to 1 for each $a \in A$, and constraint (2.10f) forces exactly one action to be selected as optimal for each contract $i \in I$. Next, we establish an equivalence between the proxy estimator PA-C and the original estimator PA.

Proposition 1. *The estimate $\boldsymbol{\pi}_n^*$ attained at a solution to PA-C is (i) a minimizer of the proxy loss function $Z_n(\boldsymbol{\pi})$, (ii) an asymptotic minimizer of the loss function $L_n(\boldsymbol{\pi})$, $|L_n(\boldsymbol{\pi}_n^*) - L_n(\hat{\boldsymbol{\pi}}_n)| \longrightarrow 0$, and (iii) consistent, $\boldsymbol{\pi}_n^* \longrightarrow \boldsymbol{\pi}^0$.*

In Proposition 1, (i) establishes that the mathematical program PA-C is equivalent to the proxy estimator (2.9), (ii) establishes that solving PA-C asymptotically produces an optimal solution to PA, and (iii) confirms that PA-C is also a consistent estimator for $\boldsymbol{\pi}^0$. Based on the equivalence in Proposition 1, we will refer to PA-C as the *exact* estimator in the remainder of the chapter.

The intuition behind Proposition 1 is as follows. Note that $Z_n(\boldsymbol{\pi})$ can be interpreted as a re-weighted version of $L_n(\boldsymbol{\pi})$, where for each (a, j), the term $|\pi_{aj} - \omega_{aj}|$ is multiplied by the weight η_{aj}/n. As $n \longrightarrow \infty$, the minimal possible loss for both estimators occurs when $\pi_{aj} = \omega_{aj}$ for all (a, j). Therefore, minimizing $Z_n(\boldsymbol{\pi})$ also minimizes $L_n(\boldsymbol{\pi})$, in the limit.

Next, note that (2.10a)–(2.10d) contains bilinear terms due to the product of the decision variables $\mathbf{x}$ and $\boldsymbol{\pi}$. Because $\mathbf{x}$ and $\boldsymbol{\pi}$ are binary and continuous variables, respectively, these product terms can be linearized exactly using well-known reformulation techniques (Glover, 1975; Adams et al., 2004), leading to a mixed-integer linear program. However, a drawback of this approach is that linearizing products of variables is known to yield weak linear programming relaxations (Adams et al., 2004; Luedtke et al., 2012), which can make solving PA-C using off-the-shelf optimization solvers challenging, even for moderately-sized data sets. In the next section, we propose an approximation to PA-C that bypasses the linearization step while remaining statistically well-

behaved.

2.4 Restricted Estimator and Statistical Column Generation

We begin this section by proposing an approximation of PA-C – which we call PA-D – based on replacing the parameter set Π with a discrete subset $\widetilde{\Pi}$ (§2.4.1). We then present a data-driven procedure for constructing the parameter set $\widetilde{\Pi}$, and investigate the behavior of the resulting estimates (§2.4.2). Then, to solve PA-D, we present a column generation algorithm based on hypothesis testing, and show that the algorithm preserves statistical consistency (§2.4.3). We conclude the section by comparing the numerical performance of the statistical column generation algorithm with off-the-shelf optimization solvers (§2.4.4).

2.4.1 Restricted estimator

Our approach to approximately solving PA-C will be to minimize the proxy loss $Z_n(\boldsymbol{\pi})$ over a restricted parameter set $\widetilde{\Pi} \subseteq \Pi$ instead of Π. The advantage of this "restricted estimator" is that the agent optimality conditions (2.10d) can be enforced without introducing bilinear terms into the formulation, which allows us to avoid the computational challenges that often accompany linearization techniques.

Next, we define a set that plays a critical role in our estimation procedure: Let $V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{|S|}\} \subseteq \mathbb{R}^d_+$ be a set of vectors indexed by S, where $\sum_{j \in J} v_{sj} = 1$ and $\mathbf{v}_s \geq 0$ for all $s \in S$. We refer to each $\mathbf{v}_s$ as a *candidate distribution*. Next, let the restricted parameter set be defined as

$$\widetilde{\Pi} = \left\{ \boldsymbol{\pi} \in \Pi \middle| \boldsymbol{\pi}_a \in V \text{ for } a \in A \right\}, \tag{2.11}$$

and let

$$\tilde{\boldsymbol{\pi}}_n = \operatorname*{argmin}_{\boldsymbol{\pi} \in \widetilde{\Pi}} Z_n(\boldsymbol{\pi}) \tag{2.12}$$

be the associated estimate. For each action $a \in A$, the parameter set $\widetilde{\Pi}$ restricts the probability distribution $\boldsymbol{\pi}_a$ to lie in the set of candidate distributions V. We assume throughout that $\widetilde{\Pi}$ is non-empty.[6]

Similar to the exact estimator (2.9), the restricted estimator (2.12) can also be formulated as a mixed-integer linear program. The intuition behind this formulation is to construct the estimate $\boldsymbol{\pi}$ in a row-wise manner by assigning a candidate distribution in V to each row of $\boldsymbol{\pi}$. To that end, let $\mathbf{w} \in \{0,1\}^{m \times S}$, $\mathbf{x} \in \{0,1\}^{n \times S}$ and $\boldsymbol{\phi} \in \{0,1\}^{n \times m \times S}$ be binary variables with the following interpretations: $w_{as} = 1$ if the candidate distribution $\mathbf{v}_s$ is assigned to be the distribution $\boldsymbol{\pi}_a$, $x_s^i = 1$ if the action assigned to candidate distribution $\mathbf{v}_s$ is optimal under contract $\mathbf{r}^i$, and $\phi_{as}^i = 1$ if the candidate distribution $\mathbf{v}_s$ is assigned to distribution $\boldsymbol{\pi}_a$ *and* action a is optimal under $\mathbf{r}^i$ and $\boldsymbol{\pi}$. Similar to PA-C, let $\mathbf{z} \in \mathbb{R}_+^{d \times S}$ be auxiliary variables used to linearize the absolute values in the loss function $Z_n(\boldsymbol{\pi})$. Then the restricted estimator (2.12) is equivalent to the following mixed-integer

[6]The parameter set $\widetilde{\Pi}$ may be empty if the requirement that $\boldsymbol{\pi}_a \in V$ for $a \in A$ conflicts with the requirement that $\boldsymbol{\pi} \in Q_\pi$ from (2.6). In this case, non-emptiness of $\widetilde{\Pi}$ can be guaranteed by projecting the candidate distributions contained in V onto the polyhedron Q_π.

linear program:

$$\underset{\boldsymbol{\pi},\mathbf{w},\mathbf{x},\mathbf{z},\phi}{\text{minimize}} \quad \sum_{s\in S}\sum_{j\in J} z_{sj} \tag{2.13a}$$

$$\text{subject to} \quad z_{sj} \geq \frac{1}{n}\sum_{i\in I}(y_j^i - v_{sj})x_s^i, \quad s \in S, j \in J, \tag{2.13b}$$

$$z_{sj} \geq \frac{1}{n}\sum_{i\in I}(v_{sj} - y_j^i)x_s^i, \quad s \in S, j \in J, \tag{2.13c}$$

$$\sum_{b\in A}\sum_{s\in S}\left(\sum_{j\in J} v_{sj}r_j^i - c_b\right)\phi_{bs}^i \geq \left(\sum_{j\in J} v_{s'j}r_j^i - c_a\right)w_{at}, \quad i \in I, a \in A, s' \in S, \tag{2.13d}$$

(PA-D)

$$\sum_{s\in S} w_{as} = 1, \quad a \in A, \tag{2.13e}$$

$$\sum_{a\in A}\sum_{s\in S}\phi_{as}^i = 1, \quad i \in I, \tag{2.13f}$$

$$x_s^i = \sum_{a\in A}\phi_{as}^i, \quad i \in I, s \in S, \tag{2.13g}$$

$$\phi_{as}^i \leq w_{as}, \quad i \in I, a \in A, s \in S, \tag{2.13h}$$

$$\pi_{aj} = \sum_{s\in S} w_{as}v_{sj}, \quad a \in A, j \in J, \tag{2.13i}$$

$$x_s^i \in \{0,1\}, \quad i \in I, s \in S, \tag{2.13j}$$

$$w_{as} \in \{0,1\}, \quad a \in A, s \in S, \tag{2.13k}$$

$$\phi_{as}^i \in \{0,1\}, \quad i \in I, a \in A, s \in S, \tag{2.13l}$$

$$\boldsymbol{\pi} \in Q_\pi. \tag{2.13m}$$

The objective (2.13a) and constraints (2.13b)–(2.13c) together represent the loss function $Z_n(\boldsymbol{\pi})$. Constraint (2.13d) enforces the agent's optimality conditions by ensuring that $\phi_{as}^i = 1$ only if candidate distribution $\mathbf{v}_s$ is mapped to $\boldsymbol{\pi}_a$, and if action a is optimal for the agent under $\mathbf{r}^i$ and $\boldsymbol{\pi}$. Constraint (2.13e) forces exactly one candidate distribution in V to be mapped to each distribution $\boldsymbol{\pi}_a$. Constraint (2.13f) ensures that only one candidate distribution in V and action $a \in A$ is selected

for contract $\mathbf{r}^i$. Constraint (2.13g) forces $x_s^i = 1$ if candidate distribution $\mathbf{v}_s$ is mapped to $\boldsymbol{\pi}_a$ and if action a is optimal under $\mathbf{r}^i$ and $\boldsymbol{\pi}$. Constraint (2.13h) ensures $\phi_{as}^i = 1$ only if $\mathbf{v}_s$ is mapped to $\boldsymbol{\pi}_a$. Constraint (2.13i) defines $\boldsymbol{\pi}_a$ as the candidate distribution from V that is assigned by $\mathbf{w}$, and constraint (2.13m) represents additional shape constraints imposed by the polyhedron Q_π. The key distinction between PA-D and PA-C is that the discrete nature of the parameter set allows the key decision variables $(\mathbf{w}, \mathbf{x}, \boldsymbol{\phi})$ to be binary, which allows us to represent the agent's optimality conditions in a way that circumvents the need for product terms.

Note that $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ represents the error in the loss function that arises from solving the restricted estimator PA-D instead of the exact estimator PA-C. Next, we present a random clustering procedure for constructing the set of candidate distributions V, and provide a finite-sample characterization of the error $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ under the proposed procedure.

2.4.2 Construction of candidate distributions and finite-sample error

Because PA-C is a consistent estimator of $\boldsymbol{\pi}^0$ (by Proposition 1), we might expect PA-D to also produce a reasonable estimate of $\boldsymbol{\pi}^0$ if the loss function error $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ is sufficiently small. Additionally, note that $Z_n(\tilde{\boldsymbol{\pi}}_n)$ is the minimal loss when the restricted parameter set $\widetilde{\Pi}$ is substituted for Π. As a result, the magnitude of the gap $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ depends on the restricted parameter set $\widetilde{\Pi}$, and by extension, the set of candidate distributions V. Here, we present a method for constructing $\widetilde{\Pi}$, based on leveraging the observed data $(\mathbf{r}^i, \xi^i)$, $i \in I$ to guide the construction of V. Our approach to constructing the candidate distributions V is summarized in Algorithm 1.

Algorithm 1 involves selecting subsets of the contract data, computing the empirical mass function over outcomes for each subset, and designating each of these empirical mass functions as a candidate distribution, $\mathbf{v}_s$. Note that the s^{th} candidate distribution is based on the outcomes of all contracts $\mathbf{r}^i$ that fall within a ball $B_s \subseteq R$; accordingly, we shall refer to the collection of data points

Algorithm 1: Sample-based construction of candidate distributions

Input: Data $(\mathbf{r}^i, \xi^i), i \in I$, parameter $\rho > 0$.
1. Randomly sample a subset S from I.
2. **for** each $s \in S$:
   $B_s = \{\mathbf{r} \in R | \|\mathbf{r}^s - \mathbf{r}\|_2 \leq \rho\}$,
   $I_s = \{i \in I | \mathbf{r}^i \in B_s\}$.
   for each $j \in J$:
   $v_{sj} = \frac{1}{n_s} \sum_{i \in I_s} y_j^i$.

Output: Candidate distributions $V = \{\mathbf{v}_s \text{ for } s \in S\}$.

indexed by I_s as the s^{th} *cluster*. The intuition for constructing the candidate distributions V in this manner is simple: Based on the agent model (1), contracts that are within a small distance of each other are likely to induce the same action from the agent. Therefore, the empirical distribution of outcomes for all contracts that lie within the ball B_s can be assumed to approximate one of the rows of the true parameter matrix $\boldsymbol{\pi}^0$ (although which row it approximates remains unknown). Next, we show that the error $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ is well behaved if V is constructed using Algorithm 1. We first impose the following assumption.

Assumption 3 (**Clustering condition**). *For each $a \in A$, there exists $s \in S$ such that $B_s \subseteq R_a(\boldsymbol{\pi}^0)$ and $I_s \neq \emptyset$.*

Assumption 3 states that for every action a, Algorithm 1 produces a ball B_s that is entirely inside the subset of the contract set R that induces action a from the agent, $R_a(\boldsymbol{\pi}^0)$. Note that if $B_s \subseteq R_a(\boldsymbol{\pi}^0)$ then every contract in cluster s induces action a from the agent. This implies that $\mathbf{v}_s$ is an empirical distribution sampled from $\boldsymbol{\pi}_a^0$. Therefore, Assumption 3 implies that for each row of $\boldsymbol{\pi}^0$, there exists at least one candidate distribution in V that is constructed by sampling from that row. Note that Assumption 3 is more likely to hold when S in Algorithm 1 is large (because we construct many balls B_s) and ρ is small (because each ball is smaller).

Our next result shows that if Assumption 3 and an additional condition on Π holds, we can

bound the approximation error $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$.

Theorem 2. *Let Assumption 3 hold, and let V be constructed using Algorithm 1. Further, suppose $\Pi = \{\boldsymbol{\pi} \geq 0 | \sum_{j \in J} \pi_{aj} = 1, a \in A\}$. Then there exists $\kappa \in (0,1)$ such that for any $\varepsilon \in (0,1)$,*

$$\Pr(|Z_n(\boldsymbol{\pi}_n^*) - Z_n(\tilde{\boldsymbol{\pi}}_n)| > \varepsilon) \leq O(n^2 \kappa^n). \tag{2.14}$$

We offer a few remarks on Theorem 2. First, observe that the bound is not monotonic due to the n^2 term, which implies that the bound can become looser in n for small n. This occurs because our proof approach depends on constructing a feasible solution $\bar{\boldsymbol{\pi}}$, and bounding the absolute number of observations where the hidden agent action is "misclassified" by $\bar{\boldsymbol{\pi}}$. Thus, the n^2 term reflects the possibility that the number of misclassified actions may increase with the sample size. Further, note that if κ is close to 1 and n is of moderate size, the bound in Theorem 2 may be vacuous. However, because it is guaranteed that $\kappa \in (0,1)$, it is straightforward to verify that $n^2\kappa^n \longrightarrow 0$, which implies that the error $Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^*)$ eventually vanishes in n.

Second, note that the rate depends on the constant κ, with lower values of κ leading to faster convergence. While κ is not particularly interpretable, it can be shown to decrease in ρ and increase in the number of clusters $|S|$. Note from Algorithm 1 that ρ is the radius of the ball B_s, for each cluster $s \in S$. Intuitively, for fixed n, larger values of ρ makes each ball B_s contain a larger number of observations, which leads to faster convergence. Conversely, larger values of $|S|$ will slow convergence, for the following reason: Because the bound depends in part on the cluster that has the fewest observations, large values of $|S|$ will increase the probability that at least one of the clusters has very few data points, which weakens the bound. Therefore, the rate $n^2\kappa^n$ is fastest when ρ is large and $|S|$ is small. However, note that Assumption 3 is more likely to hold in the opposite case: when ρ is small and $|S|$ is large. Therefore, selecting ρ and $|S|$ requires balancing

their effects on κ with ensuring that Assumption 3 holds.

Third, observe that the bound expression is invariant to ε provided $\varepsilon \in (0, 1)$. Intuitively, this occurs because the key object of interest in the proof is a sequence of Bernoulli variables (which contribute to the loss function error in a binary manner) that we use to bound the number of times the hidden action is misclassified by a constructed solution $\bar{\boldsymbol{\pi}}$. However, we note that ε does indeed appear in the non-dominant terms of the bound, as we would expect (see (A.38) in the proof of Theorem 2).

Note that Theorem 2 is only valid for the case where each $\boldsymbol{\pi}_a$ is permitted to be any valid probability vector (i.e., $Q_\pi = \mathbb{R}^{m \times d}$). This additional condition is imposed on Π because the randomness of the set V can render the solution constructed by our proof approach infeasible for a more general parameter set Π. However, this additional assumption on Π is only needed for the finite-sample characterization of the error in Theorem 2; Proposition 2 below shows that the solution from the restricted estimator, $\tilde{\boldsymbol{\pi}}_n$, is asymptotically optimal with respect to the exact estimator PA-C for any Π that satisfies (2.6).

Proposition 2. *Let Assumption 3 hold. Then PA-D is asymptotically optimal with respect to PA-C:* $|Z_n(\boldsymbol{\pi}_n^*) - Z_n(\tilde{\boldsymbol{\pi}}_n)| \longrightarrow 0$.

The asymptotic optimality established in Proposition 2 provides assurance that PA-D is a reasonable approximation to PA-C when n is large, which is precisely the regime where PA-C is likely to be intractable. As a consequence, we should also expect the restricted estimator to produce "good" estimates of $\boldsymbol{\pi}^0$ for larger sample sizes. Having established that PA-D reasonably approximates PA-C, we now focus on developing a solution technique for tackling the mixed-integer program PA-D.

2.4.3 Statistical column generation

Observe that the size of the optimization problem PA-D grows with the number of candidate distributions in V, which can make PA-D computationally intractable if V is large. In this section, we propose a solution algorithm that involves solving PA-D over a subset of V – which we shall call V^+ – which dramatically improves the tractability of the estimator PA-D, with minimal degradation in estimation error. Because each candidate distribution in V is mapped to a set of decision variables in PA-D (where the set S indexes the distributions in V), our solution technique can be interpreted as a column generation algorithm.

The key step of our approach is a series of non-parametric hypothesis tests, which identifies a subset V^+ by performing pairwise comparisons of candidate distributions in V. The intuition is as follows. Consider any candidate distribution $\mathbf{v}_s \in V$, and recall from Algorithm 1 that $\mathbf{v}_s$ is the empirical mass function over outcomes associated with the contracts in the s^{th} cluster. If there exists another cluster s' such that all contracts in clusters s and s' induce the same action from the agent, then $\mathbf{v}_s$ and $\mathbf{v}_{s'}$ can be interpreted as two empirical mass functions that were generated by the same probability distribution (i.e., one of the rows of $\boldsymbol{\pi}^0$). Therefore, our goal will be to apply non-parametric hypothesis tests to identify whether any pairs in V are generated by the same distribution, and to discard those that are effectively "duplicates".

Hypothesis test function.

A hypothesis test typically consists of four main steps: (1) a null hypothesis is specified that we wish to test, (2) a significance level α (i.e. Type I error rate) is specified for the test, (3) a test statistic is computed based on the sample data, and (4) the null hypothesis is rejected if and only if the magnitude of the test statistic exceeds a threshold τ_α, where τ_α depends on α. In the context of our column generation algorithm, the null hypothesis we will test is whether two candidate

distributions $\mathbf{v}_s$ and $\mathbf{v}_{s'}$ are generated from the same probability distribution (i.e., the same $\boldsymbol{\pi}_a^0$), for many pairs (s, s').

We first introduce some additional definitions that are required by our algorithm. For each $s \in S$, define a vector $\boldsymbol{\psi}_s \in \mathbb{Z}_+^d$, where the j^{th} entry is the frequency of outcome j in the s^{th} cluster of Algorithm 1. The vector $\boldsymbol{\psi}_s$ is simply a convenient form for representing the candidate distribution $\mathbf{v}_s$ within our hypothesis tests. Let $n_s = |I_s|$ be the number of observations in cluster s, and note $n_s = \sum_{j \in J} \psi_{sj}$ for $s \in S$. Next, note that for each $s \in S$, by the weak law of large numbers there exists $\boldsymbol{\nu}_s \in \mathbb{R}_+^d$ such that $\|\boldsymbol{\nu}_s - \mathbf{v}_s\| \longrightarrow 0$ as $n_s \longrightarrow \infty$. We now define the main ingredient of the algorithm, which is a *test function* that declares whether $\boldsymbol{\psi}_s$ and $\boldsymbol{\psi}_{s'}$ are statistically different at a significance level α.

Definition 1. *$H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) : \mathbb{Z}_+^d \times \mathbb{Z}_+^d \mapsto \mathbb{R}$ is a test function if $\Pr(H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) > 0 | \boldsymbol{\nu}_s \neq \boldsymbol{\nu}_{s'}) \longrightarrow 1$ as $n_s \longrightarrow \infty$ and $n_{s'} \longrightarrow \infty$ and $\Pr\left(H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) > 0 | \boldsymbol{\nu}_s = \boldsymbol{\nu}_{s'}\right) \leq \alpha$.*

Definition 1 states that the hypothesis test function returns a positive value if and only if the null hypothesis – that the candidate distributions $\mathbf{v}_s$ and $\mathbf{v}_{s'}$ are generated by the same probability distribution – is rejected. This definition subsumes many two-sample, non-parametric hypothesis tests. One example is the Kolmogorov–Smirnov hypothesis test (Massey, 1951; Stephens, 1974), which is widely used for its ease of implementation. In particular, the test function is given by

$$H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) = \sup_{j \in J} \left| \frac{\psi_{sj}}{n_s} - \frac{\psi_{s'j}}{n_{s'}} \right| - \tau_\alpha \sqrt{\frac{n_s + n_{s'}}{n_s n_{s'}}},$$

where τ_α^{KS} is the critical value associated with a significance level of α (Smirnov, 1948). We note here that the Kolmogorov-Smirnov test is known to be conservative for discrete distributions (Slakter (1965); Conover (1972)). As a result, selecting τ_α based on Kolmogorov–Smirnov critical values for continuous distributions makes α an upper bound on the true Type I error rate in

our setting, but otherwise does not affect the validity of our algorithm. Other examples of non-parametric tests that fit within our framework are the Anderson-Darling (Anderson et al., 1952; Scholz and Stephens, 1987), Chi-squared (Cochran, 1952) and the Cramér–von Mises (Anderson, 1962) tests.

Algorithm overview.

Let S^+ index the candidate distributions in V^+. We let PA-D(S^+) denote formulation PA-D where S is replaced with the subset S^+, and we let PA-D(S) denote the original formulation with the full set V. Let $V^- = V \setminus V^+$ and $S^- = S \setminus S^+$ denote the omitted distributions and the accompanying index set, respectively. Given a significance level α, we shall say two candidate distributions $\mathbf{v}_s$ and $\mathbf{v}_{s'}$ are *statistically different* if and only if $H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) > 0$; that is, the null hypothesis that $\mathbf{v}_s$ and $\mathbf{v}_{s'}$ were generated from a common probability distribution is rejected. In each iteration of the main loop of the algorithm, we perform a series of hypothesis tests identify a new candidate distribution to be introduced to V^+, and solve PA-D(S^+) once there does not exist any distribution in V^- that is statistically different from every distribution in V^+ at a significance level of α. Specifically, in each iteration we compute

$$s^* = \operatorname*{argmax}_{s \in S^-} \inf_{s' \in S^+} H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}).$$

Intuitively, $\mathbf{v}_{s^*}$ is the distribution in V^- that is the "most" different from all distributions in V^+, based on the selected test function H_α. The distribution $\mathbf{v}_{s^*}$ is then added to V^+ if and only if

$$\inf_{s' \in S^+} H_\alpha(\boldsymbol{\psi}_{s^*}, \boldsymbol{\psi}_{s'}) > 0. \tag{2.15}$$

If (2.15) holds, then $\mathbf{v}_{s^*}$ is statistically different from every distribution in V^+, and is thus added to V^+. If (2.15) does not hold, then there are no remaining distributions in V^- that are statistically different from all distributions in V^+. In this case, we solve PA-D(S^+), and the algorithm terminates. A summary is given in Algorithm 2.

Algorithm 2: Statistical column generation (PA-D+)

Input: Data $(\mathbf{r}^i, \xi^i)$, $i \in I$, candidate distributions V produced by Algorithm 1, significance level $\alpha > 0$.
Initialize: Set $t = 0$. Select any $s \in S$. Set $S^+ = \{s\}$ and $S^- = S \setminus \{s\}$.
1. Let $s^* = \text{argmax}_{s \in S^-} \inf_{s' \in S^+} H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'})$.
if $\inf_{s' \in S^+} H_\alpha(\boldsymbol{\psi}_{s^*}, \boldsymbol{\psi}_{s'}) \leq 0$ or $S^- = \emptyset$,
Solve PA-D(S^+) and obtain solution $\boldsymbol{\pi}_n^+$, set $T = t$, and **terminate**.
else Update $t \leftarrow t + 1$, $S^+ \leftarrow \{S^+, s^*\}$, and $S^- \leftarrow S^- \setminus \{s^*\}$. Return to Step 1.
Output: Parameter estimate $\boldsymbol{\pi}_n^+$, iteration count T.

We will use "PA-D+" to denote the estimator represented by Algorithm 2. There are two main differences between existing column generation methods for large-scale integer programs and the one we propose in Algorithm 2. First, the column generation process in Algorithm 2 involves performing several hypothesis tests, which are fast to compute. By comparison, existing methods for integer programs typically generate columns by solving an auxiliary optimization problem (often called the *pricing* problem due to its use of dual information), which is often an integer program itself and may be difficult to solve (Lubbecke and Desrosiers, 2005). Second, Algorithm 2 is not guaranteed to produce an optimal solution to PA-D; in contrast, the purpose of existing column generation methods is to solve the "original" optimization problem exactly. Therefore, Algorithm 2 effectively sacrifices optimality for computational efficiency. However, although Algorithm 2 does not produce optimal solutions to PA-D, it can be shown to produce a consistent estimate of $\boldsymbol{\pi}^0$, which is our main objective in this work.

Consistency and iteration bound.

Next, we present the main result of §2.4: Theorem 3 shows that the approximate solution obtained by Algorithm 2 preserves the consistency of the exact estimator PA-C.

Theorem 3. *Let $\boldsymbol{\pi}_n^+$ be the estimate obtained by PA-D+ (Algorithm 2). Then*

$$\boldsymbol{\pi}_n^+ \longrightarrow \boldsymbol{\pi}^0.$$

As a consequence of Theorem 3, we should expect $\boldsymbol{\pi}_n^+$ to provide a reasonable estimate of the unknown parameter $\boldsymbol{\pi}^0$. However, note that Theorem 3 is an asymptotic result only, and that for small n the estimate from PA-D+ may be less accurate than the exact estimate obtained by solving PA-C. We compare the performance of these two approaches numerically in §2.4.4.

Note that because the termination condition in Algorithm 2 depends on the outcome of a series of hypothesis tests, the total number of iterations, denoted by T, is a random variable. In Theorem 4 below, we show that $\mathbb{E}[T]$ is bounded by a function of the problem parameters, including the significance level α used in the hypothesis testing step of Algorithm 2.

Theorem 4. *Let Assumption 3 hold. Further, assume that for each $s \in S$, there exists $a \in A$ such that $B_s \subseteq R_a(\boldsymbol{\pi}^0)$. Then*

$$\mathbb{E}[T] \leq m[1 + \alpha \cdot |S| \cdot (|S| - m)].$$

The proof of Theorem 4 relies on upper bounding $\Pr(T > m)$ – the probability that the number of iterations in Algorithm 2 exceeds the number of agent actions. In particular, we show in the proof of Theorem 4 that $\Pr(T > m) \leq \alpha m S$. The intuition for the preceding inequality is as follows. Observe that by construction, the candidate distribution v_s is the empirical distribution over outcomes associated with all contracts $\mathbf{r}^i$ such that $\mathbf{r}^i \in B_s$. Because for each $s \in S$, $B_s \subset$

$R_a(\boldsymbol{\pi}^0)$ for some $a \in A$ (by assumption), there are at most m unique distributions from which the empirical distributions v_s are generated, which are $\boldsymbol{\pi}_a^0$, $a \in A$. Next, note that in Algorithm 2, a candidate distribution is only added to the set V^+ if the hypothesis testing step finds it to be statistically different from every distribution in V^+. Therefore, the event $\{T > m\}$ implies that a Type I error has occurred at some point during Algorithm 2; that is, a candidate distribution was added to $V+$ despite the underlying distribution $\boldsymbol{\pi}_a^0$ already being "represented" in $V+$ by another candidate distribution.

Because α bounds the probability of making a Type I error, smaller values of α will make Algorithm 2 more *conservative* in adding new distributions to V^+, thus increasing the probability of the event $\{T > m\}$. Conversely, if α is large, then it becomes more likely that a given distribution v_s is determined to be statistically different from those in V^+, which leads to more distributions being added to V^+, and thus a greater number of iterations. The dependence on S arises for a similar reason; as S increases, so does the number of omitted distributions V^-, which increases the likelihood that there exists a distribution in V^- that satisfies the inclusion criterion in Step 2 of Algorithm 2.

Additionally, note that the bound $\mathbb{E}[T] \leq |S|$ holds trivially, because $T = |S|$ implies $S_T^- = \emptyset$ by Algorithm 2. As a result, the bound in Theorem 4 may be vacuous if α is large, but is made meaningful for an appropriate selection of S and α. It is also straightforward to verify that the assumption in the statement of Theorem 4 implies that $|S| \geq m$, which confirms that the bound on $\mathbb{E}[T]$ is strictly positive for all $\alpha > 0$.

2.4.4 Numerical performance

In this section, we compare the performance of three estimation methods using synthetic data. The first two are solving the exact estimator (PA-C) and the restricted estimator (PA-D) directly with

optimization software. The third is solving the restricted estimator using the column generation technique outlined in Algorithm 2 (PA-D+). We focus our comparison on the solution times and estimation errors from the three approaches.

Setup.

Recall that m and d denote the number of actions and outcomes, respectively. We consider five problem sizes, given by $(m, d) \in \{(2, 2), (4, 5), (5, 10), (10, 20), (20, 40)\}$. For each of the five problem sizes, we consider three sample sizes, given by $n \in \{100, 500, 1000\}$. Then for each combination (m, d, n), we randomly generate $\boldsymbol{\pi}^0$ from the appropriately-sized parameter set Π given by (2.6), where Q_π is given by (2.7). For each (m, d, n), we randomly generate contract data by sampling $\mathbf{r}^i$ uniformly from $[1, 10]^d$ for each $i = 1, \ldots, n$, and sampling $\mathbf{c}$ uniformly from $[0, 1]^m$. The outcome associated with each $\mathbf{r}^i$ is obtained by solving the agent's problem (2.1) under the corresponding $\boldsymbol{\pi}^0$. We repeat this procedure for a total of 10 trials for each (m, d, n). To parameterize PA-D, we set $S = 50$ and $\rho = 10 \times d$. For the hypothesis testing step for PA-D+, we use the discrete analogue of the two-sample Anderson–Darling test (Scholz and Stephens, 1987), and set $S = 50$, $\rho = 10 \times d$ and $\alpha = 0.05$. We use the optimization solver Gurobi 8.0 to solve PA-C, PA-D and PA-D+.

Results.

Table 2.1 summarizes the average solution time and estimation errors over 10 trials for the three estimators. In each trial, the error associated with PA-C, PA-D and PA-D+ is given by $\frac{1}{md}\|\boldsymbol{\pi}^0 - \boldsymbol{\pi}_n^*\|$, $\frac{1}{md}\|\boldsymbol{\pi}^0 - \tilde{\boldsymbol{\pi}}_n\|$, and $\frac{1}{md}\|\boldsymbol{\pi}^0 - \boldsymbol{\pi}_n^+\|$, respectively. In all trials, we set a time limit of 3600 CPU seconds. Dashes in the table indicate instances where an optimal solution was not found within 3600 CPU seconds for any of the 10 trials. In many of these trials, no feasible solution was found within 3600 CPU seconds; we therefore only include errors obtained at optimal solutions to PA-C or PA-D when

reporting the average estimation error.

We offer a few observations regarding Table 2.1. First, note that for each problem size, the estimation error generally decreases in n, which corroborates our consistency results (Proposition 1 and Theorem 3, respectively). Second, for smaller problem instances (e.g., $m = 4$, $d = 5$, $n = 1000$), PA-C is less computationally expensive than PA-D+, which we posit is a consequence of requiring fewer binary decision variables. However, PA-D+ generally scales more efficiently in the problem and sample size than PA-C and PA-D, with the most notable performance improvement occurring at larger problem instances (e.g., $m = 10$, $d = 20$, $n = 1000$). Third, solving the restricted estimator PA-D directly with Gurobi is less tractable than solving the exact estimator PA-C with Gurobi. This is again likely attributable to PA-D requiring many more binary variables than PA-C, due to how the restricted parameter set is represented in the formulation PA-D. Nonetheless, the results indicate that this intractability can be overcome by (approximately) solving the restricted estimator using the statistical column generation technique, without significantly compromising estimation error. Fourth, observe that larger problem sizes are not necessarily more computationally expensive – for example, the average solution time of the instances $(2, 2, 1000)$ and $(5, 10, 1000)$ for PA-C was 245 and 12 seconds, respectively. We conjecture that this is because the larger problem sizes offer the estimator additional degrees of freedom in fitting the agent model to the data (due to containing a larger number of unknown parameters), which allows the optimization problem to more quickly attain the minimal objective function value. Lastly, the favorable performance of PA-D+ in the larger instances (e.g., $m = 20$, $d = 40$) suggests that our estimator and algorithm can also be used to tractably approximate contracting problems with continuous actions and outcomes through discretization.

Note that the purpose of Algorithm 2 is not to generate a provably optimal solution to PA-D, which is typically the case with similar column generation methods. Instead, our primary goal is

			PA-C		PA-D		PA-D+	
m	d	n	Time	Error	Time	Error	Time	Error
2	2	100	2	0.07	20	0.06	2	0.09
2	2	500	19	0.06	3432	0.06	4	0.06
2	2	1000	245	0.06	–	–	15	0.06
4	5	100	0	0.05	–	–	4	0.09
4	5	500	2	0.05	–	–	18	0.06
4	5	1000	3	0.05	–	–	66	0.06
5	10	100	1	0.04	–	–	4	0.06
5	10	500	6	0.03	–	–	14	0.04
5	10	1000	12	0.03	–	–	47	0.03
10	20	100	2404	0.02	–	–	3	0.02
10	20	500	–	–	–	–	15	0.01
10	20	1000	–	–	–	–	26	0.01
20	40	100	–	–	–	–	2	0.01
20	40	500	–	–	–	–	84	0.01
20	40	1000	–	–	–	–	211	0.01

Table 2.1: Solution time (CPU seconds) and normalized estimation error of three formulations averaged over 10 trials. Instances that did not solve to optimality under 3600 CPU seconds are omitted when calculating average estimation error. Dashes indicate no instance solved to optimality within 3600 CPU seconds in any trial.

to generate an estimate of the true parameter $\boldsymbol{\pi}^0$ that is statistically consistent, competitive with solutions from solving the exact estimator, and attainable in a computationally efficient manner. Theorem 3 and the numerical results in Table 2.1 suggest that Algorithm 2 meets each of these criteria.

2.5 Empirical Study: Randomizing Incentives in a Crowdwork Platform

In this section, we demonstrate our method by using it to investigate the effect of financial incentives on work quality in an online labor platform. First, we conducted an experiment on a crowdwork platform (Amazon Mechanical Turk) by randomly assigning incentive contracts to a pool of workers completing the same task. We then estimate an agent model from the experimental data, which

allows us to characterize the link between incentives and quality and solve for an optimal incentive contract.

2.5.1 Background: Incentives and quality on Amazon Mechanical Turk

Crowdwork platforms are used by businesses that require temporary labor to complete tasks that are typically difficult for computers but simple for humans. Common tasks include audio transcription, classification of images, and data entry. The largest and most well-known crowdwork platform is Amazon's Mechanical Turk ("mTurk"), which has been estimated to have 100,000 unique workers, with 2,000 active at any given time (Difallah et al., 2018).

The mTurk platform allows "requesters" to post tasks, along with a reward to be paid to the worker upon successful completion. Workers can select the tasks they want to complete, typically on a first-come, first-served basis. Requesters have discretion over whether to pay workers for their submissions, and can deny payment if the worker's submission is incomplete or low quality. Requesters can also provide bonuses to workers. Workers can be informed of the structure of the bonus payment within the instructions for a task, which offers the requester considerable flexibility in designing incentives.

The question of whether financial incentives improve quality of work in crowdwork platforms has been addressed in multiple studies, with differing conclusions. Mason and Watts (2009) find that incentives improve the quantity, but not quality of work; similarly, Yin et al. (2013) find that the magnitude of the bonus does not affect quality. In contrast, Horton and Chilton (2010) and Harris (2011) both find that quality can improve with worker pay. An important study in this line of research is by Ho et al. (2015), who suggest that for tasks where quality plausibly depends on worker effort (e.g., proofreading), incentives can improve quality.

With respect to experimental design, we underline two differences between our study and the

work cited above. First, instead of assigning workers to a finite number of treatments (e.g., bonus or no bonus), we vary incentives in a continuous manner, meaning the parameters of the incentive contract are randomly drawn *for each worker*. This design significantly complicates the implementation of the experiment on mTurk, but introduces useful variation for estimating our agent model. Second, we examine how incentives affect the *distribution* of work quality, instead of average quality.

2.5.2 Experimental setup

Task design.

A major source of observable heterogeneity in the mTurk worker population is location. Approximately 91% of workers are located in two countries: the US (75%) and India (16%) (Difallah et al., 2018). We collected and analyzed data from both countries separately.

The experiment involved posting two types of tasks on mTurk. First, we posted a recruitment task in which workers were paid $1.00 for agreeing to be notified of future tasks by email. We recruited 250 workers from both the US and India using this task, for a total worker pool of 500. The recruitment task in each country was made available for one day, and reached its maximum number of submissions (250 for each location) within 3 hours of posting. Second, inspired by Ho et al. (2015), we created a proofreading task by inserting 10 typos into a one-page, 500-word excerpt from a newspaper article. The proofreading task required workers to report the line number and correct spelling for each misspelled word in the article (e.g., "5:automobile"). We use a proof reading task because it allows us to objectively measure the quality of each submission (percentage of typos identified). After constructing the worker pool, we posted the proofreading task on mTurk and notified each worker by email of the task's availability. The task was available for 24 hours.

Incentive structure and randomization.

We next describe how we randomized incentives among workers. The mTurk platform allows requesters to assign "qualification" criteria to tasks, which only allows workers with the required qualifications to view and complete the task. For example, a requester might assign a location or age-based qualification to a task if they wish to target a specific worker population. Requesters can also create and assign custom qualifications to workers. When conducting a randomized experiment, creating and randomly assigning qualifications to workers effectively allows the requester to construct multiple treatment groups, where each qualification represents one treatment.

We use the qualification feature in mTurk to create exogenous variation in worker incentives. We first created 500 unique qualifications, and assigned each qualification to a single worker in the pool. We then created 500 tasks where each task was randomly assigned to a qualification. As a result, for each of the 500 tasks, only a single worker in our pool was able to view and complete it.

The payment for the proofreading task consisted of two components: a *base* payment for finding at least 25% of the typos in the document, and an additional *bonus* payment for finding at least 75% of typos. For each task (i.e., each worker in the pool), we drew *base* and *bonus* uniformly from the interval [$0.10, $1.00], rounded to the nearest $0.01. We provided the details of the payment structure upfront in the task instructions. Because workers were only able to view the task assigned to them, workers could not observe the payment offered to others, and had no knowledge that payments were randomized. Note that in the context of the proofreading task, worker output corresponds to the fraction of typos corrected, which we also refer to as the task *quality*.

Submissions.

We collected a total of 346 submissions, each from a unique mTurk worker. Of these, 215 submissions were from US-based workers, and 131 were from India-based workers. We analyze the data

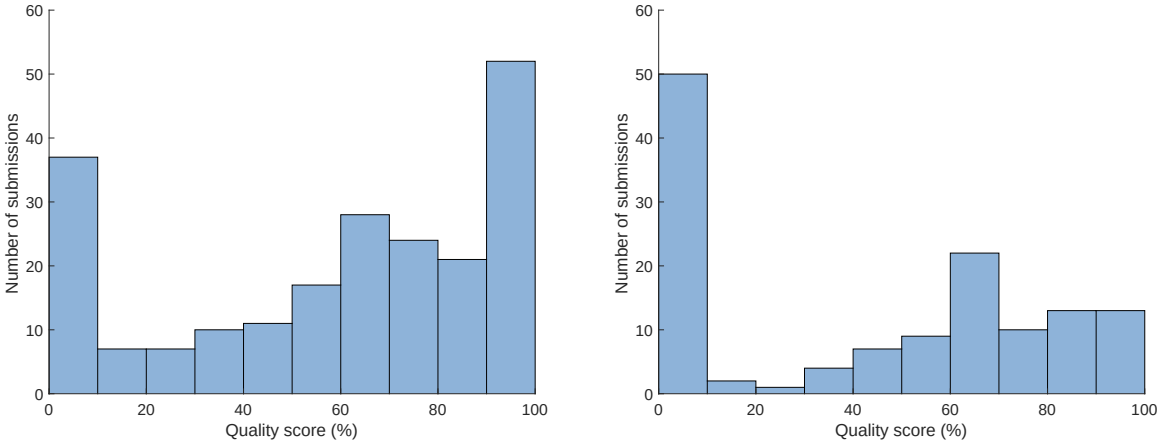
from US and India workers separately throughout our study. Figure 2.1 depicts the distribution of quality scores for workers in each location. Note that a large number of submissions achieve a quality score of 0. Low-quality submissions are a well-known feature of mTurk; because verifying responses manually for a large number of submissions is difficult, workers may submit blank or low-quality responses in the hope of nevertheless receiving a payment (Ipeirotis et al., 2010). Scores of 0 may also be due to submissions not being in the correct format, which we specified as a condition for payment in the task instructions.

The mTurk platform provides timestamps for when a worker accepted and submitted a task. The mean completion time (i.e., time between acceptance and submission) was 9.7 minutes, and 95% of workers submitted the task between 1 and 29 minutes after accepting it. Because mTurk allows workers to accept tasks into a queue before working on them, the recorded completion time is an upward-biased measurement of the actual time the worker spent on the task. As a result, completion time may be a poor proxy for true worker effort, because the requester cannot observe how much time the task spent in the worker's queue. We therefore treat effort as fully hidden, and do not use completion time data in our study.

Based on each worker's completion time, we estimated the average wage to be \$14.50/hr for our task (including the guaranteed \$1.00 payment at recruitment). This is likely a conservative estimate of the true average wage due to the queueing behavior described above.

2.5.3 Estimation and validation

Next, we describe the application of our estimation procedure to the experimental data. Putting the results of the experiment in the format required by our estimator is straightforward. Recall that each worker was eligible for three possible payments based on their submission quality: no payment (if they found 0-25% of typos), a base payment (25%-75%), or both a base and bonus



(a) US. (b) India.

Figure 2.1: Distribution of quality scores for submissions made by workers in the US and India.

payment (75-100%). In our framework, this corresponds to $d = 3$ possible performance levels for the worker's outcome ξ^i. Accordingly, the i^{th} worker's incentive contract $\mathbf{r}^i$ has the components $r_1^i = 0$, $r_2^i = base^i$, and $r_3^i = base^i + bonus^i$, where $base^i$ and $bonus^i$ are the randomly generated parameters for that worker. For the PA-D+ algorithm, we set $\rho = 0.5$, $S = 10$ and $\alpha = 0.0001$ throughout all experiments.

Measuring goodness of fit.

We require a goodness of fit metric for fitting and validating the model. Recall that our estimation procedure generates a prediction of the outcome distribution: Given an estimate $\hat{\boldsymbol{\pi}}$, a contract $\mathbf{r}$, and action costs $\mathbf{c}$, the model's prediction of the outcome distribution under $\mathbf{r}$ is $\hat{\boldsymbol{\pi}}_{a(\mathbf{r})}$, where $a(\mathbf{r})$ is the agent's optimal action under contract $\mathbf{r}$. For ease of interpretation, we measure goodness of fit as the absolute error between the empirical and predicted probability of a given outcome, averaged over all outcomes. Specifically, let $(\mathbf{r}^i, \xi^i)$, $i = 1, \ldots, n$ be the data we wish to measure our model fit against. As before, for each i, let $y_j^i = 1$ if $\xi^i = j$ (if outcome j is observed). Then

the mean absolute error (MAE) is given by

$$\text{MAE} = \frac{1}{d}\sum_{j=1}^{d}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\pi}_{a(\mathbf{r}^i),j} - y_j^i\right)\right|. \tag{2.16}$$

Setting cost parameters.

Two hyperparameters in our model are the number of actions, m, and the action costs, $c_1, \ldots, c_m$. We selected these parameters using a standard 10-fold cross-validation procedure, using MAE to measure cross-validation errors. To avoid performing an extremely large number of cross-validation iterations, we imposed additional structure by assuming action costs were of the form $c_a = (a-1)\cdot\delta$, for $a = 1, \ldots, m$. We used cross-validation to jointly select m and δ from the sets $\{2, 3, 4\}$ and $\{0.02, 0.05, 0.1, 0.2, 0.5\}$, respectively (units of the latter set are dollars). Results are presented in Table 2.2. Note that errors are relatively stable for all values of δ when $m = 2$ or $m = 3$, whereas the model appears to overfit for $m = 4$. We select $(m, \delta) = (3, 0.1)$ for both the US and India datasets, resulting in the cost vector $\mathbf{c} = [0,\ 0.1,\ 0.2]$. Lastly, for Algorithms 1 and 2, we set $\rho = 0.5$, $S = 50$, and $\alpha = 0.01$, and use a Chi-squared test for the hypothesis test step of Algorithm 2.

We note that our handling of action costs is fairly stylized, because they are treated as hyperparameters to be tuned prior to model fitting. Horton and Chilton (2010) estimates the median reservation wage of mTurk workers to be \$1.38/hr. Given that the median completion time for our task was 8 minutes, \$0.00–\$0.20 appears to be a reasonable approximation for the range of effort costs of an mTurk worker.

		δ				
	m	0.02	0.05	0.1	0.2	0.5
US	2	0.06	0.06	0.07	0.06	0.06
	3	0.04	0.06	0.04	0.06	0.05
	4	0.18	0.11	0.12	0.12	0.08
IN	2	0.07	0.08	0.06	0.06	0.08
	3	0.07	0.05	0.06	0.06	0.08
	4	0.15	0.08	0.06	0.11	0.09

Table 2.2: 10-fold cross-validation errors (MAE) for US and India groups, with varying number of actions (m) and cost spacing (δ).

Bootstrapping.

Given our moderately-sized data set ($n = 215$ and $n = 131$), we validated our estimation procedure by bootstrapping. For each of 100 repetitions, we sampled n observations with replacement, and estimated the model parameters from the sample using Algorithm 2. For each repetition, we assessed model fit using two hypothesis tests: a Chi-squared (χ^2) test, which is appropriate in our setting because outcomes are discrete, and an exact test using MAE as the test-statistic, where the sampling distribution is obtained through Monte Carlo simulation. In both hypothesis tests, the null hypothesis is that the empirical distribution of quality outcomes in the out-of-bootstrap data is generated by the fitted model. Accordingly, we interpret large p-values as indicating a good model fit.

Table 2.3 shows the distribution of test statistics and associated p-values over the 100 bootstrap repetitions. Both tests produced comparable p-values within each worker group. Note that the median p-value was above 0.1 for both groups, which suggests the model reasonably fits the joint distribution over $(\mathbf{r}, \xi)$ in the majority of bootstrap iterations.

Table 2.4 presents the estimated values of $\boldsymbol{\pi}$ and standard errors for both worker groups. Each 3×3 section in the center of Table 2.4 corresponds to the estimated $\boldsymbol{\pi}$ matrix for the labelled worker group, averaged over 100 bootstrap repetitions. For convenience, we refer to the outcome

		5^{th}	25^{th}	Median	75^{th}	95^{th}
US	χ^2 (p-value)	0.20 (0.90)	1.70 (0.43)	3.42 (0.18)	7.36 (0.03)	23.58 (0.00)
	MAE (p-value)	0.02 (0.89)	0.04 (0.42)	0.07 (0.14)	0.10 (0.02)	0.15 (0.00)
IN	χ^2 (p-value)	0.18 (0.91)	0.99 (0.61)	2.13 (0.34)	5.23 (0.07)	14.33 (0.00)
	MAE (p-value)	0.02 (0.93)	0.05 (0.65)	0.09 (0.35)	0.14 (0.09)	0.19 (0.00)

Table 2.3: Percentiles of Chi-squared (ξ^2) and MAE test statistics with associated p-values over 100 bootstrap repetitions.

	Actions (a)	Outcomes (j)			# Obs.
		1	2	3	
US	1	0.46 (0.18)	0.34 (0.18)	0.20 (0.09)	19
	2	0.30 (0.10)	0.42 (0.12)	0.28 (0.08)	27
	3	0.20 (0.07)	0.43 (0.07)	0.37 (0.06)	169
IN	1	0.57 (0.12)	0.34 (0.12)	0.09 (0.07)	36
	2	0.45 (0.09)	0.38 (0.11)	0.17 (0.06)	36
	3	0.35 (0.07)	0.42 (0.07)	0.23 (0.06)	59

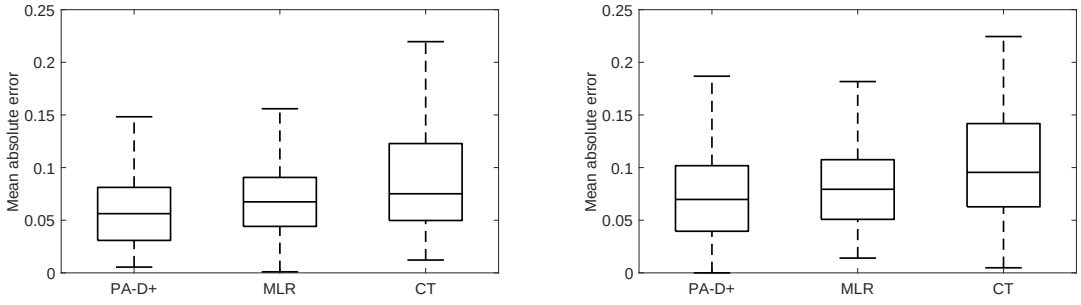
Table 2.4: Estimated values of $\boldsymbol{\pi}$ for both groups, with standard errors in parentheses. The final column reports the number of in-bootstrap observations mapped to each action, averaged over 100 bootstrap repetitions.

in which the worker earns the bonus ($\xi^i = 3$) as the "bonus outcome", and the probability that this outcome is realized as the "bonus probability". Note that the highest cost action ($a = 3$) has the highest bonus probability in both worker groups, and that the bonus probability is lower in the India worker group compared to the US group, for all actions.

Our estimation procedure treats each action a as a latent variable. The solution to the estimation problem produces a clustering where each outcome is assumed to have been generated by one of the m distributions (i.e., agent actions). As a result, for each bootstrap repetition, we can count the number of observations that are assigned to each action by the estimator. The average number of observations mapped to each action are reported in the final column of Table 2.4.

Predictive performance.

Next, we evaluate the predictive performance of the estimator. For each of the 100 bootstrap models, we compute the prediction error (given in (2.16)) attained by the fitted model on the out-of-bootstrap observations. We set $S = 10$, $\rho = 0.5$, and $\alpha = 0.0001$. To serve as performance benchmarks, we repeat the bootstrap procedure for standard implementations of multinomial logistic regression (MLR) and classification trees (CT), both of which also generate predictions of the outcome distribution for a given set of contracts.[7] Figure 2.2 depicts the distribution over prediction errors for the three methods over the 100 bootstrap repetitions. For the US data, the average MAE for PA-D+, MLR and CT is 0.059, 0.070 and 0.093, respectively; for the India data, the average errors are 0.072, 0.086, and 0.112. In summary, Figure 2.2 confirms that the PA-D+ estimator produces sound predictions on the experimental mTurk data, and is competitive with well-known benchmark methods. In §A.2 of the electronic companion, we further compare all three methods on several synthetic instances, and find that our estimator continues to perform well.



(a) US data.

(b) India data.

Figure 2.2: Comparison of out-of-bootstrap prediction errors for PA-D+, multinomial logistic regression (MLR) and classification trees (CT) on mTurk data (100 repetitions).

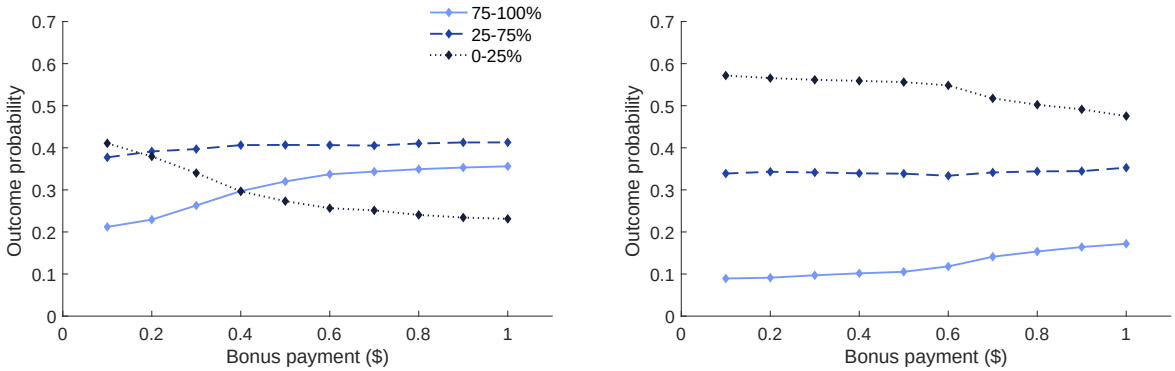
[7]Both benchmark methods are implemented using MATLAB's Statistics and Machine Learning Toolbox using default settings.

2.5.4 Impact of bonuses on quality

We now use the estimated model to examine the effect of varying the bonus payment on quality. For a given incentive contract, we form a prediction of the outcome distribution by averaging over the 100 bootstrapped models, which improves stability and reduces overfitting (Breiman, 1996). Let $\hat{\boldsymbol{\pi}}^1, \ldots, \hat{\boldsymbol{\pi}}^K$ be the estimates obtained from K bootstrap repetitions. The probability of observing outcome j under the incentive contract $\mathbf{r}$ is then given by $\frac{1}{K}\sum_{k=1}^{K} \hat{\boldsymbol{\pi}}^k_{a^k(\mathbf{r}),j}$, where $a^k(\mathbf{r})$ is the optimal action for contract $\mathbf{r}$ in the k^{th} agent model. To isolate the influence of the bonus payment, we fix the base payment to \$0.10, vary the bonus payment between \$0.10 and \$1.00, and compute the probability of each quality outcome under each bonus amount. We repeat for a base payment of \$1.00.

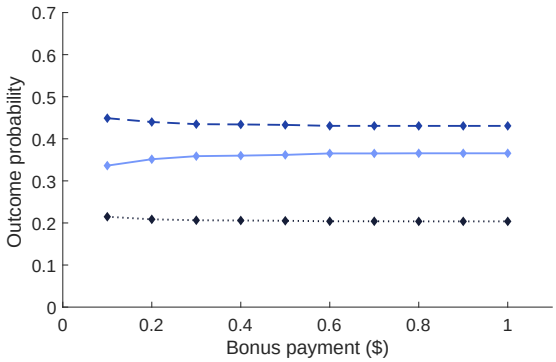
Figure 2.3 shows the results for both the US and India groups of workers. For a base payment of \$0.10 (Figure 2.3a and 2.3b), the bonus probability (i.e., probability that submission quality is above 75%) increases moderately for both groups as the bonus is increased from \$0.10 to 1.00 (from 0.21 to 0.36 for the US group; from 0.09 to 0.17 for the India group). However, with a base payment of \$1.00 (Figure 2.3c and 2.3d), the effect of increasing the bonus payment from \$0.10 to \$1.00 is dampened (bonus probability increases from 0.34 to 0.37 for the US group; 0.18 to 0.22 for the India group). These results suggest that increasing the bonus payment can indeed increase quality, but the effect is significantly diminished when the base payment is already high. A qualitatively similar result can be obtained by fixing the bonus payment and varying the base payment (results not shown).

We shed some light on the mechanics behind Figure 2.3. Because our predictions are based on the average of 100 different agent models, for a fixed incentive contract, we can count the number of models in which each action is taken. Further, note that if the bonus payment increases, an agent may find it optimal to "switch" from a low-cost action to a high-cost action, thus increasing the

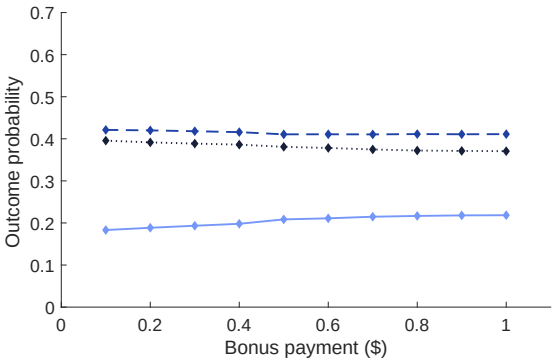


(a) US, base = \$0.10.

(b) India, base = \$0.10.



(c) US, base = \$1.00.



(d) India, base = \$1.00.

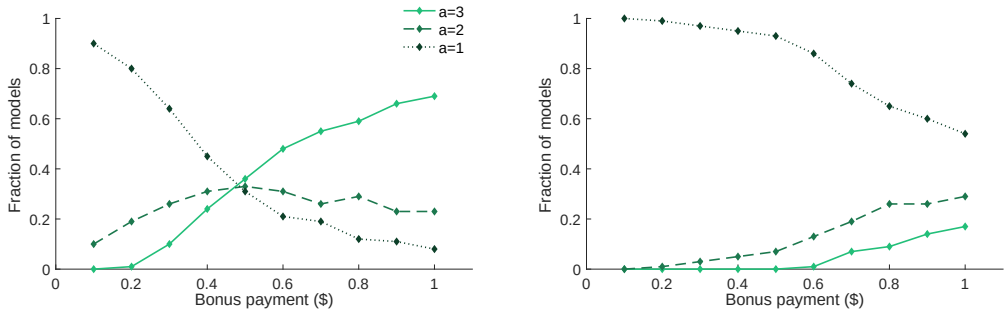
Figure 2.3: Effect of varying bonus payment on probability of each quality outcome (0-25%, 25-75%, 75-100%) for US and India workers.

probability of realizing the high-quality outcome. The change in probabilities depicted in Figure 2.3 is the result of the underlying agent models jumping from one action to the next as the parameters of the contract change.

Figure 2.4 shows the fraction of agent models that take each of the three actions as the bonus is increased from \$0.10 to \$1.00. Note that the four panels in Figure 2.4 map to the four panels in Figure 2.3. As expected, when the base payment is \$0.10, increasing the bonus amount from \$0.10 to \$1.00 is associated with agents switching away from the lowest cost action ($a = 1$) toward the higher cost actions ($a = 2$ and $a = 3$). Moreover, the shift toward higher cost actions is more pronounced for the US worker group, where the fraction of agents taking the highest cost action ($a = 3$) increases from 0 to 0.69; for the India group, this fraction increases from 0 to 0.17. In parallel with Figure 2.3, when the base payment is \$1.00, the fraction of agents taking the highest cost action ($a = 3$) is higher overall, but the shift toward higher cost actions as the bonus is increased is muted. In other words, the stability in selected actions shown in Figures 2.4c and 2.4d explains the stability in outcome probabilities seen in Figures 2.3c and 2.3d. We emphasize here that Figure 2.4 is intended to illustrate the mechanics behind the predictions in Figure 2.3, and is not necessarily a depiction of worker behavior.

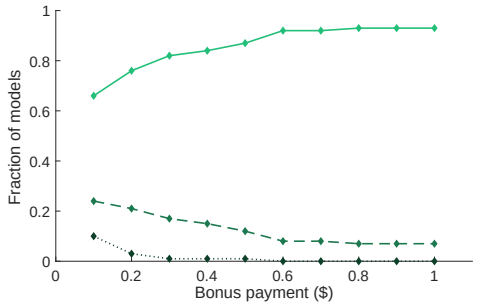
2.5.5 Solving for an optimal incentive contract

An advantage of our model specification is that it leads to an optimal contracting problem that is highly tractable (see §A.1 of the electronic companion for details). To illustrate this in the context of our mTurk study, we consider the simple problem of maximizing the bonus probability (i.e.,

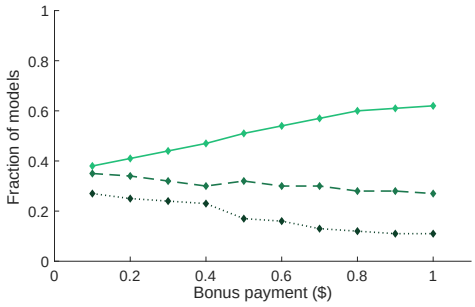


(a) US, base = \$0.10.

(b) India, base = \$0.10.



(c) US, base = \$1.00.



(d) India, base = \$1.00.

Figure 2.4: Effect of bonus payment on optimal agent actions in 100 bootstrapped models.

outcome $\{\xi = 3\}$) subject to a budget constraint on the expected payment:

$$\underset{\mathbf{r}}{\text{maximize}} \quad \hat{\pi}_{\hat{a}(\mathbf{r}),3} \tag{2.17a}$$

$$\text{subject to} \quad \hat{a}(\mathbf{r}) = \underset{a \in A}{\text{argmax}}\, \mathbf{r}^\top \hat{\boldsymbol{\pi}}_a - c_a, \tag{2.17b}$$

$$\mathbf{r}^\top \hat{\boldsymbol{\pi}}_{\hat{a}(\mathbf{r})} \leq \Gamma, \tag{2.17c}$$

$$\mathbf{r} \geq \mathbf{0}. \tag{2.17d}$$

The formulation above is a special case of the general optimal contracting problem presented in §A.1 of the electronic companion, and can be solved exactly by solving $|A|$ linear programs. An important consequence of the tractability of the optimal contracting problem (2.17) is that we can easily characterize the performance of the optimal contracts as the budget parameter Γ varies. To

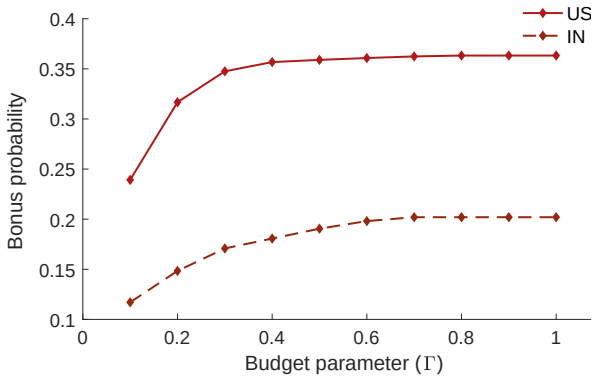


Figure 2.5: Frontier of optimal bonus probabilities under varying budget parameter Γ.

do so, we solve (2.17) for each $\Gamma \in \{0.05, 0.1, \ldots, 1\}$ (for each of the 100 bootstrap estimates), and compute the average bonus probability under each value of Γ.

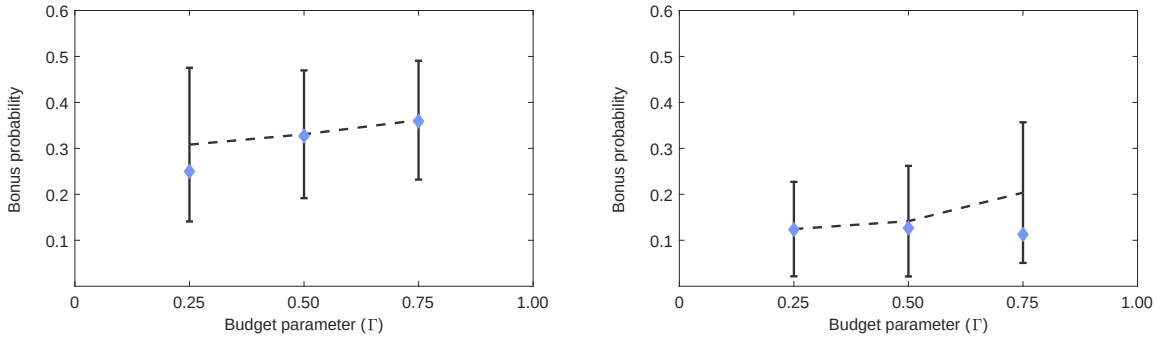
Figure 2.5 shows the resulting frontiers for both the US and India worker data. Because the curves are obtained by solving the optimal contracting problem (2.17), they represent estimates of the maximum attainable performance for both worker groups over the entire class of contracts used in the experiment. The value of the budget parameter Γ can be interpreted as the expected payment to the agent under the corresponding optimal contract. Our main finding is that higher payments increase quality modestly: increasing the expected payment from \$0.10 to \$1.00 increases the bonus probability under the optimal contract by 0.08–0.12, depending on the worker group. However, the most striking observation is that returns to quality diminish at fairly low payment levels, with quality improvements leveling off around \$0.30 and \$0.60 for the US and India groups, respectively.

Figure 2.5 also clearly depicts the difference in the performance of optimal contracts between the US and India worker groups. For example, for the US group, attaining a bonus probability of 0.30 requires an expected payment of at least \$0.20; for the India group, a bonus probability of 0.30 is not attainable through higher payments alone. It can also be observed that the bonus probability is approximately 0.10–0.15 higher among US workers across all payment levels.

2.5.6 Experimental validation of contract performance

To validate the predicted performance of the optimal contracts shown in Figure 2.5, we conducted six follow-up experiments on mTurk. First, for each of the 100 bootstrap estimates $\hat{\boldsymbol{\pi}}^1, \ldots, \hat{\boldsymbol{\pi}}^K$, we solved the optimal contracting problem (2.17) for $\Gamma \in \{0.25, 0.50, 0.75\}$, which corresponds to three different points on the frontiers in Figure 2.5. We then computed the optimal contract by taking the component-wise average of the 100 solutions to (2.17). This produced six different testable contracts (i.e., combinations of the *base* and *bonus* parameters), which are shown in Table 2.5. We implemented each contract on mTurk by recruiting a new pool of 600 unique workers (using the same approach described in §2.5.2), and assigning 100 workers to each of the six contracts. Table 2.5 summarizes the results from these experiments, including the empirical bonus probability for each contract (i.e., the fraction of submissions with quality above 75%). In Figure 2.6, we plot the empirical bonus probabilities along with the 95% prediction intervals obtained from the bootstrap.

Figure 2.6 shows that for each of the six experimentally tested contracts, the empirical bonus probability sits comfortably inside its corresponding prediction interval, and is often close to the midpoint of the interval. In general, the prediction intervals are wide, which is unsurprising given that many other factors likely influence submission quality beyond the payment amount, including unobserved worker attributes. Further, validating the predictions from any model through experiments is challenging in general; because the worker population on mTurk is not temporally static (Difallah et al., 2018), the worker population in the validation experiments may be different from the initial experiments used to estimate the model. Nevertheless, our results in Figure 2.6 suggest that the estimator can reasonably predict experimental outcomes under a given incentive contract.



(a) US. (b) India.

Figure 2.6: Empirical bonus probabilities and 95% prediction intervals of six contracts implemented on mTurk.

	Budget (Γ)	Base	Bonus	Submissions	$> 75\%$	Empirical Bonus Probability
US	0.25	0.16	0.42	20	5	0.20
	0.50	0.23	0.47	52	17	0.33
	0.75	0.41	0.78	39	14	0.36
IN	0.25	0.11	0.31	73	9	0.12
	0.50	0.25	0.42	79	10	0.13
	0.75	0.58	0.77	71	8	0.11

Table 2.5: Optimal incentive contracts under three different values of Γ and associated results from mTurk experiments.

2.5.7 Discussion

Our results suggest larger incentives can increase quality on crowdwork platforms, corroborating the results of Ho et al. (2015). While similar results are reported in the literature, we have taken a complementary approach by characterizing worker performance over a *class* of incentive contracts. Further, the tractability of the optimal contracting problem under our agent model allows us to estimate performance under an optimal contract. In particular, as summarized in Figure 2.5, we find that increasing the expected worker payment by about \$1 increases the probability that a worker crosses the bonus threshold by 0.08–0.12, depending on the worker's location. Most notably, we find diminishing returns to quality at relatively low payments in both worker groups, which may help explain why requesters tend to set low wages on mTurk (Hara et al., 2018).

We also observe that quality can depend strongly on the worker's location. In particular, as seen in Figure 2.5, the bonus probability for the India group at an expected payment \$1.00 is comparable to the US group at \$0.10. This result aligns with a finding by Shaw et al. (2011), who observe that quality on mTurk is much more strongly associated with worker location than financial incentives. While we have only focused on worker location in this study, our approach can be readily extended to other worker attributes, provided sufficient data is available.

We highlight some limitations of our study and note directions for future work. First, we have treated agent costs as hyperparameters by tuning them through cross-validation. This makes the costs used in our model a rough approximation of actual worker costs, and may limit the interpretability of the resulting agent model. Our agent model also does not capture many of the worker dynamics present in crowdwork platforms. Horton and Chilton (2010) point out that mTurk worker output appears to deviate from what would be predicted by simple, rational agent models, which applies to our model as well. Lastly, an important aspect of crowdwork not addressed here is worker welfare. In particular, mTurk has been widely criticized for low worker pay, which is often far below the US minimum wage (Hara et al., 2018). While we did not address worker welfare in this work, our modeling framework can also be used to characterize welfare over a class of incentive contracts, and allows for welfare considerations to be explicitly incorporated into the optimal contracting problem (e.g., by imposing constraints on agent utility). Investigating the trade-off between worker welfare and quality in crowdwork may be a fruitful direction for future work.

2.6 Conclusion

We proposed an approach for estimating parameters that govern agent production in a moral-hazard principal-agent model. First, we presented an estimator for a non-parametric agent model,

and showed it to be statistically consistent. To avoid computational drawbacks of solving the estimator exactly, we proposed an approximate estimator based on a restricted parameter set, and characterized the approximation error both asymptotically and in a finite-sample setting. To solve the restricted estimator, we developed a novel column generation technique that uses hypothesis testing to select variables, which we showed preserves consistency. Numerical results show that the approximation scheme and solution technique produce accurate estimates in a computationally efficient manner. Lastly, we applied our estimator to data from a randomized experiment on a crowdwork platform to demonstrate how our method can be used to characterize performance over a class of incentive contracts and identify optimal incentives from the estimated model.

We conclude by noting some possible directions for future work. Our estimation procedure is built upon a general but simple moral-hazard agent model; it may be useful to extend our approach to accommodate other common features of principal-agent models, such as unobserved heterogeneity and risk aversion. There may also be fertile ground in generalizing our statistical column generation algorithm to other integer programming problems. In particular, our approach may be relevant to other estimation problems where the parameter space is a very large set of discrete distributions. Lastly, estimating an agent model from data may be valuable for investigating questions related to worker welfare, which is an issue of increasing prominence in online labor platforms.

Chapter 3

Discovering Causal Models with Optimization

3.1 Introduction

The availability of high quality data has made integrating predictive and prescriptive models increasingly promising. These models often consist of two components: a statistical procedure for learning parameters of interest from observational data, and a prescriptive model that takes the estimated parameters as input. This combination of methodologies has been successful in improving decision-making in a variety of contexts, including supply chain management, revenue management, and healthcare (Giesecke et al., 2018; Mišić and Perakis, 2020).

An implicit assumption of the paradigm described above is that the decision maker has knowledge of the *causal* relationship between observed variables – that is, that the fitted statistical model alone is sufficient for predicting the effects of a hypothetical decision. However, if the underlying causal assumptions are incorrect, then the model can yield misleading or even harmful prescriptions (Kallus and Zhou, 2020). In recognition of the importance of causality, there is a recent but limited body of work that explicitly incorporates causal inference into prescriptive models (Bertsimas and Kallus, 2016; Alley et al., 2019; Kallus and Zhou, 2020; Baardman et al., 2020; Gupta et al., 2020). While these approaches address the challenge of estimating causal effects when the variables are

subject to confounding, they primarily focus on narrowly circumscribed causal structures over the variables of interest.

In this chapter, we take a step back and consider how to learn causal relations from data in the first place using the framework of graphical causal models. The framework of graphical causal models (Spirtes et al., 2000; Pearl, 2000), discussed in more detail in §3.2, enables this sort of inference by providing a precise mathematical representation of the causal structure of a system (in terms of a directed graph) and the observed data (in terms of a probability distribution associated with the graph). Using this framework, a variety of causal discovery methods have been developed to infer underlying causal structures from observational data. With a few important exceptions, these methods have relied on two restrictive assumptions, which limit their practical relevance. The first is the absence of latent confounders – referred to as *causal sufficiency* – which means that there are no unmeasured common causes of the measured variables. The second is that there is no feedback, meaning the causal structures can be represented by directed *acyclic* graphs (DAGs). Both of these assumptions can only rarely be justified in practice.

Our main contribution is a new optimization-based method for causal discovery that allows for both unmeasured confounders *and* feedback cycles. To address the scalability issues resulted from considering this general search space, we propose an iterative solution technique that exploits the conditional (in)dependence structure in the data to detect "promising" candidate edges in the underlying graph, which are then assembled into a causal graph by an optimization model. The generality of our method combined with its computational efficiency greatly expands the practical relevance of causal discovery to empirical research.

The remainder of the chapter is organized as follows. §3.2 reviews foundational concepts in causal discovery, which a familiar reader may skip. §3.3 presents an optimization model for causal discovery with latent confounders and feedback cycles. §3.4 develops a solution technique for the

model. §3.5 extends our framework to address the challenge of feature selection. §3.6 concludes with a brief discussion of further developments.

3.2 Causal Graphical Models

In the framework of causal graphical models (Pearl, 2000; Spirtes et al., 2000), causal relations are represented by a directed graph $\mathcal{G} = (V, E)$ over a set of nodes V, where the edge set E contains directed edges, representing a direct (relative to V) causal effect of one node on another. We will first introduce the theory and notation using *directed acyclic graphs* (DAGs), as they permit the most intuitive explanation and the simplest causal interpretation (§3.2.1). We then extend these concepts to include graphs that contain cycles and unobserved confounding variables (§3.2.2).

3.2.1 Acyclic Models Without Unmeasured Confounding

In acyclic causal models without unmeasured confounding, the graph $\mathcal{G} = (V, E)$ over a set of nodes V contains at most one directed edge between any pair of nodes. We define an edge $e \in E$ as a triple (i, t, j) with $i, j \in V$, $i \neq j$, and $t \in \{\rightarrow, \leftarrow\}$. Central to our algorithm is the notion of a *path* between two variables:

Definition 2 (Path)**.** *Given a node set V, a set of edge-types $T = \{\rightarrow, \leftarrow, \leftrightarrow\}$ and an edge set E of triples (v_1, t, v_2) with $v_1, v_2 \in V$ and $t \in T$, we define a path p_{ij} from node i to node j with $i, j \in V, i \neq j$, as a sequence of edges $p_{ij} = (e_1, \ldots, e_\ell)$ such that $e_k \in E$ for all $1 \leq k \leq \ell$, e_1 starts with node i, e_ℓ ends with node j, consecutive edges are connected, and nodes on the path do not repeat (other than as start- and endpoint of consecutive edges). See Appendix B.1 for formalization.*

A *directed path* from i to j is then a path where all edges point towards j. Any node connected by a directed path from i is a *descendant* of i, any node connected by a directed path to i is an *ancestor* of i. *Parents* and *children* of a node i are the direct causes and effects, respectively, of i

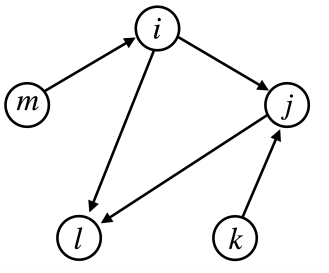


Figure 3.1: Example directed acyclic graph (DAG).

in $\mathcal{G}$. A DAG is a directed graph in which there is no pair of distinct nodes (i, j) such that there is a directed path from i to j and an edge $j \to i$. We say that a node i is a *collider* on a path if its adjacent edges point into i ($\to i \leftarrow$). A *non-collider* on a path is a node i that is either a *mediator* ($\to i \to$) or a *common cause* ($\leftarrow i \to$). For example, in Figure 3.1, node j is a collider on the path $i \to j \leftarrow k$ and a mediator on the path $i \to j \to l$, and node i is a common cause on the path $l \leftarrow i \to j$.

In causal modeling, a DAG $\mathcal{G}$ is associated with a probability distribution $P_{\mathcal{G}}(V)$, which describes causal relations over the set of nodes V.[1] A standard assumption is that the distribution is generated by the graph structure such that it factorizes: $P_{\mathcal{G}}(V) = \prod_{i \in V} P_{\mathcal{G}}(i|Pa(i))$, where $Pa(i)$ are the parents of node i in $\mathcal{G}$ (Spirtes and Zhang, 2016; Eberhardt, 2017). Based on the connection between the causal structure and the resulting data distribution, many causal discovery algorithms, including the one we present here, exploit the independence structure seen in the data to infer the underlying causal relations. One of the central concepts required for this inference is the notion of *d-separation* (Geiger et al., 1990), which can be thought of as the graphical counterpart to probabilistic independence. It is based on the notion of a *blocked path*:

Definition 3 (Blocked paths)**.** *A path between nodes i and j is* unblocked *with respect to a set of nodes C if every collider k on the path is in C or has a descendant in C, and no other nodes on*

[1] Given the correspondence between graphical structure and probability distribution, we will use the terms "node" and "variable" interchangeably.

the path are in C. Otherwise the path is blocked *with respect to C (Pearl, 2000).*

Definition 4 (d-separation)**.** *Two nodes i and j are* d-separated *with respect to a conditioning set C (denoted by $i \perp j|C$) if all paths between them are blocked, otherwise they are* d-connected *given C (denoted by $i \not\perp j|C$) (Pearl, 2000).*

To illustrate the definitions above, note that there are two paths between m and l in Figure 3.1: $m \to i \to l$ and $m \to i \to j \to l$. By Definition 3, both of these paths are unblocked with respect to the empty conditioning set $C = \{\}$, which by Definition 4 implies that m and l are d-connected with respect to $C = \{\}$. Now consider the conditioning set $C = \{i\}$. By Definition 4, conditioning on node i blocks these paths because node i is a noncollider on both of these paths. Since there does not exist an unblocked path between nodes m and l with respect to the conditioning set $C = \{i\}$, it follows that m and l are d-separated with respect to the conditioning set $C = \{i\}$.

With this definition in hand, we can now put (conditional) d-separation in correspondence with (conditional) independence. Following convention, we use the single turnstile ($\perp$) to denote d-separation and the double turnstile ($\perp\!\!\!\perp$) to denote probabilistic independence. There are two standard assumptions that are used to achieve this correspondence: the *causal Markov* condition and the *faithfulness* condition.

Assumption 4 (Causal Markov)**.** *If node i is d-separated from node j given conditioning set C in graph $\mathcal{G} = (V, E)$ with $i, j \in V$ and $C \subseteq V \setminus \{i, j\}$, then i is probabilistically independent of j given C in the distribution over the graph $P_{\mathcal{G}}(V)$:*

$$i \perp j|C \text{ in } \mathcal{G} \quad \Longrightarrow \quad i \perp\!\!\!\perp j|C \text{ in } P_{\mathcal{G}}(V). \tag{3.1}$$

Assumption 5 (Faithfulness)**.** *If variable i is probabilistically independent of variable j given conditioning set C in the distribution over the graph $P_{\mathcal{G}}(V)$, then i is d-separated from j given C*

in graph $\mathcal{G} = (V, E)$*:*

$$i \perp\!\!\!\perp j | C \text{ in } P_{\mathcal{G}}(V) \quad \Longrightarrow \quad i \perp j | C \text{ in } \mathcal{G}. \tag{3.2}$$

The (global) causal Markov condition, as we have stated it here, follows from how we have defined the probability distribution in terms of the causal structure (Pearl, 2000). In contrast, the faithfulness condition, which is the converse to the Markov condition, represents an additional assumption, as it ensures that an independence in the data is actually due to a d-separation (rather than, for example, two causal pathways cancelling each other out (Spirtes et al., 2000; Uhler et al., 2013)). Together, the causal Markov and faithfulness conditions provide a tight correspondence between (conditional) probabilistic independence and (conditional) d-separation.

Remark 1. *Under the causal Markov and faithfulness conditions, a (conditional) independence in* $P_{\mathcal{G}}(V)$ *is present if and only if there is a corresponding (conditional) d-separation in DAG* $\mathcal{G}$*.*

This correspondence is the basis for many causal discovery methods, as one can now use the independence structure in the data to constrain the graph structure.

3.2.2 Extension to Cyclic Models with Latent Confounding

We introduced the key concepts in the context of directed acyclic graphs (DAGs). In the remainder of this chapter, we focus on a more general class of graphs that permit cycles and can represent confounding due to unobserved variables. For such graphs, many of the key ideas above can be generalized, but they require much more book-keeping and are much less intuitive. We briefly outline the required adjustments here.

A cyclic model, as its name suggests, permits feedback cycles in the causal structure. In this setting, the edge set E in $\mathcal{G} = (V, E)$ may contain an edge in each direction between a pair of nodes.[2]

[2]In principle there can also be edges from a node to itself, but such self-loops are redundant for linear Gaussian

These cycles do not represent backwards-in-time causation, but should instead be understood as shorthand notation for causal feedback that is unravelled over time; for example, $i_t \to j_{t+1} \to i_{t+2}$. One of the simplest and most well-studied cyclic causal models is the linear Gaussian cyclic model given by $\mathbf{x}(t) = \mathbf{B}\mathbf{x}(t-1) + \epsilon$, where $\mathbf{x}$ is a vector representation of the variables in V, ϵ is a vector of independent errors, and $\mathbf{B}$ is a square matrix representing the (possibly cyclic) causal effects among the variables (Hyttinen et al., 2012). Under appropriate conditions, the model converges to an equilibrium, which allows cycles such as $i_t \to j_{t+1} \to i_{t+2}$ to be represented more simply without time indices as $i \overrightarrow{\leftarrow} j$.[3] For this type of linear Gaussian cyclic model, the Markov and faithfulness conditions still imply the correspondence between (conditional) d-separation and (conditional) independence (Remark 1). However, unlike the acyclic case, the correspondence between d-separation and probabilistic independence does not hold in general in cyclic models.

Remark 2. *In the cyclic case, the correspondence between d-separation and probabilistic independence holds for linear Gaussian causal models, but not in general (Spirtes, 1995).*

We will restrict consideration of *cyclic* models to the linear Gaussian case in order to utilize the correspondence described in Remark 1.[4] In the *acyclic* case, our results are not restricted to a particular parameterization.

To represent confounding between a pair of variables (i, j) due to an unobserved common cause c, the graphical framework is extended to include the bi-directed edge $i \leftrightarrow j$ (see Figure 3.2(b) and (c)). Here, $i \leftrightarrow j$ is shorthand for $i \leftarrow c \to j$, where c is an unobserved variable ($c \notin V$). The graph $\mathcal{G} = (V, E)$ then consists of a set of variables V and edges E such that every pair (i, j) is permitted to contain directed edges $(\to, \leftarrow)$, possibly in both directions, and a bi-directed edge

cyclic models that we consider here (Hyttinen et al., 2012).

[3] In the general case with arbitrary initial conditions, a sufficient and necessary condition for convergence to an equilibrium is for all eigenvalues of $\mathbf{B}$ to be less than 1 (Hyttinen et al., 2012).

[4] An extension to more general parameterizations for cyclic models is beyond the scope of this work – see the notion of σ-separation in Forré and Mooij (2018) for a thorough treatment.

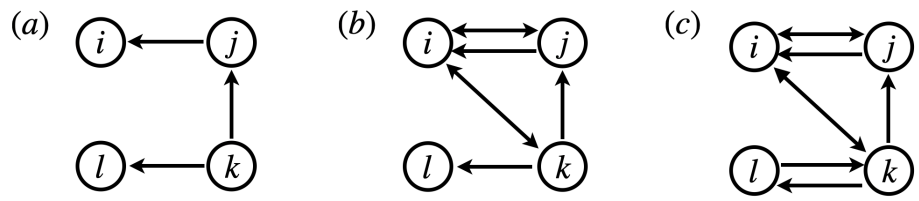


Figure 3.2: Example graphs: (a) A directed acyclic graph (DAG). (b) The more general acyclic directed mixed graph (ADMG), which does not contain cycles, but contains bi-directed edges to represent unmeasured confounding. (c) A directed mixed graphs (DMG) that allows for both cycles and confounding.

($\leftrightarrow$) to represent unmeasured confounding.[5] Colliders remain defined as before, but now they can also arise from bi-directed edges incident on the "colliding" variable.

The model class that includes bi-directed edges but disallows cycles is often referred to as *acyclic directed mixed graphs* (ADMGs) (Figure 3.2(b)). Our focus will be on the general model class of *directed mixed graphs* (DMGs), where both bi-directed edges and cycles are allowed (Figure 3.2(c)). D-separation can be naturally extended to DMGs, one just has to keep track of a larger set of possible edges, since any pair of variables can now be connected by three different edge types. In the cyclic linear Gaussian model described above, confounding can be represented using a non-diagonal covariance matrix for the error terms, resulting in correlated errors. For such linear Gaussian models with correlated errors the correspondence between d-separation and probabilistic independence (Remark 2) still holds (Spirtes, 1995).

3.2.3 Constraint-Based Causal Discovery

Our proposed method belongs to a class of causal discovery algorithms known as *constraint-based* methods (see Maathuis et al. (2010), Spirtes and Zhang (2016) and Eberhardt (2017) for reviews). These methods involve performing conditional independence tests on the data to construct conditional (in)dependence "constraints", that are used to search for a causal graph that satisfies these

[5] Only one bi-directed edge is used to represent all possible confounders between a pair of variables. A confounder of n observed variables is represented by $\binom{n}{2}$ bi-directed edges among the n variables.

constraints to the extent possible. Here, each input constraint is the statement $i \perp\!\!\!\perp j|C$ or $i \not\perp\!\!\!\perp j|C$ for some $i, j \in V$ and $C \subset V \setminus \{i, j\}$, which implies a d-separation or d-connection that the output graph must satisfy (by Remark 1).

One of the first and most well-known constraint-based methods is the PC-algorithm by Spirtes et al. (2000), which is restricted to searching for the equivalence classes of DAGs. Alternative methods generalize the search space to acyclic graphs with unmeasured common causes (Spirtes et al., 2000) and cyclic causal models (albeit without unmeasured confounding) (Richardson, 1996). A number of variants of these methods have been developed in the literature with the aim of improving computational efficiency or reliability (Colombo et al., 2012; Teramoto et al., 2014). Constraint-based methods have been shown to be asymptotically correct for their respective background assumptions, meaning in the large-sample limit they discover the true data-generating graph up to an equivalence class (see §3.3.2 for details) (Spirtes et al., 2000; Zhang and Spirtes, 2002; Solus et al., 2017).

An advantage of constraint-based causal discovery is that it allows the user to completely separate the statistical challenge of establishing the (conditional) independence constraints from the combinatorial inference of finding graphs that are consistent with them. Thus, one can choose independence tests that are suitable for the particular domain and adopt their preferred correction method for multiple hypothesis testing. Such decisions will be informed by the sample size, the number and dimensionality of the variables, whether the variables are categorical or continuous, or what assumptions one is willing to make about the parametric form of the causal relations.

Our work is most closely related to constraint-based methods that allow for both cycles and unmeasured confounders. These methods encode d-separation constraints obtained from independence tests in a logical representation, and either use Boolean satisfiability solvers (Hyttinen et al., 2013), *answer set* solvers (Hyttinen et al., 2014, 2017) or custom branch-and-bound algorithms

(Rantanen et al., 2018, 2020) to identify a graph that minimizes the (weighted) sum of violated d-separation constraints. Because focusing on DMGs dramatically expands the search space of possible graphs, the discovery task can be computationally challenging for all of these methods, even when the number of variables is modest (i.e., fewer than 10).

To overcome limited scalability, other causal discovery methods have considered simplifications such as (i) only searching for causal ancestry relations, rather than direct causal connections (Magliacane et al., 2016)), (ii) only allowing unmeasured confounders, but no cycles (Triantafillou and Tsamardinos, 2015), or (iii) by weakening the faithfulness assumption (Zhalama et al., 2017). Instead of imposing restrictive assumptions, our approach maintains tractability by iteratively expanding the search space of possible graphs, and optimizing the solution using two alternating integer programs.

There are a small number of existing methods for causal discovery that are explicitly based on integer optimization (Jaakkola et al., 2010; Cussens, 2012; Bartlett and Cussens, 2017; Park and Klabjan, 2017; Kucukyavuz et al., 2020; Manzour et al., 2021). These methods all focus on DAGs, and thus do not accommodate feedback cycles or unobserved confounders. A second distinction is in the formulation of the problem – ours is a constraint-based approach, and accordingly searches for a graph that satisfies conditional (in)dependencies seen in the data. In contrast, the papers cited above present *score-based* methods, which typically involves maximizing the likelihood of the data under a given DAG. Such a formulation of the search problem is not easily generalized to handle cyclic and confounded models.

3.3 Path-Based Model for Causal Discovery

Our model takes as input a set of (conditional) independence and dependence relations over the variables V. Assuming the Markov and faithfulness conditions (and for the cyclic case, that the

parameterization is linear Gaussian), these relations imply a corresponding set of d-separation and d-connection relations that must be satisfied by the output graph (Remarks 1 and 2). At a high level, our model conceptualizes the search space of DMGs as combinations of paths, and aims to select a set of paths such that the resulting graph minimizes violations of the input d-separation and d-connection constraints.

3.3.1 Model Formulation

Given a data set over a set of variables V, we denote by $C \subset V$ a generic set of conditioning variables. We then define $A_{ij} = \{C \mid C \subseteq V \setminus \{i, j\}\}$ to be the set of all possible conditioning sets C for the pair (i, j). We refer to the n^{th} such conditioning set in A_{ij} as $C_{ij}^n \in A_{ij}$. Then, let $D_{ij} = \{C \in A_{ij} \mid i \not\perp\!\!\!\perp j|C\}$ be the set of all conditioning sets such that i and j are statistically *dependent* conditional on C; and similarly, $I_{ij} = \{C \in A_{ij} \mid i \perp\!\!\!\perp j|C\}$ to be the set of all conditioning sets such that i and j are statistically *independent* conditional on C. We assume that each pair of variables (i, j) is either dependent or independent given a particular conditioning set C. Thus, for all pairs (i, j) we have $D_{ij} \cup I_{ij} = A_{ij}$ and $D_{ij} \cap I_{ij} = \emptyset$. We use N_{ij}, N_{ij}^D and N_{ij}^I as index sets for A_{ij}, D_{ij} and I_{ij}, respectively, where $N_{ij} = N_{ij}^D \cup N_{ij}^I$.

Based on the equivalence established in Remark 1, the sets D_{ij} and I_{ij} encode the d-separation and d-connection relations found in the data. A graph $\mathcal{G}$ satisfies the d-connection implied by $C \in D_{ij}$ if $i \not\perp_{\mathcal{G}} j|C$; similarly, $\mathcal{G}$ satisfies the d-separation implied by $C \in I_{ij}$ if $i \perp_{\mathcal{G}} j|C$. Our objective is to find a directed mixed graph $\mathcal{G}$ that minimizes the number of d-separation and d-connection constraints found in the data that are not satisfied in $\mathcal{G}$; that is, we want to find a graph $\mathcal{G} \in \mathcal{O}$ where

$$\mathcal{O} = \left\{ \mathcal{G} \,\middle|\, \mathcal{G} \in \arg\min_{\mathcal{G}} \left(\sum_{i,j \in V} \sum_{C \in D_{ij}} \mathbf{1}(i \perp_{\mathcal{G}} j|C) + \sum_{i,j \in V} \sum_{C \in I_{ij}} \mathbf{1}(i \not\perp_{\mathcal{G}} j|C) \right) \right\}. \tag{3.3}$$

We formulate the problem of searching for a graph in $\mathcal{O}$ as an integer program (IP). Our model constructs a graph by selecting *paths* that best fit the discovered d-connection and d-separation conditions. Despite the possible cycles in DMGs, we show in Appendix B.1 that checking only a finite number of finite length paths is sufficient to determine all d-connection relations. [6]

Let $P_{ij}^{-}(\tilde{E})$ be the set of all paths between nodes i and j over a given edge set $\tilde{E}$ as defined by Definition 2. Let col_p be the set of colliders on path p. We then define an *appendage of* p, to capture possible d-connections obtained by conditioning on descendants of colliders, to be an (acyclic) directed path that one can construct with edges in $\tilde{E}$ that has as root a collider in col_p and does not pass through i or j. For each path p we can then generate a set of *extended paths* $P_{pij}^{+}(\tilde{E})$ that contains p and all combinations of p with its appendages. (See Appendix B.1 for details) Finally, we define the set of all extended paths between i and j given an edge set $\tilde{E}$ as $P_{ij}(\tilde{E}) = \bigcup_{p \in P_{ij}^{-}} P_{pij}^{+}(\tilde{E})$ and the set of all extended paths given an edge set $\tilde{E}$ as $\mathcal{P}(\tilde{E}) = \bigcup_{\{(i,j) \in V : i \neq j\}} P_{ij}(\tilde{E})$.

We define a path's *length* ℓ_p as the unique number of edges in the sequence. For each $p \in \mathcal{P}(\tilde{E})$ and $e \in \tilde{E}$, let ϕ_{pe} be a parameter where $\phi_{pe} = 1$ if and only if edge e belongs to path p, and let α_{ijp}^{n} be a parameter where $\alpha_{ijp}^{n} = 1$ if and only if the path $p \in P_{ij}(\tilde{E})$ is *unblocked* with respect to the conditioning set $C_{ij}^{n} \in A_{ij}$.

Next, we define three types of binary decision variables. Let $\mathbf{x} \in \{0,1\}^{|\tilde{E}|}$ determine edges in the graph $\mathcal{G} = (V, E)$ with $E \subseteq \tilde{E}$, where $x_e = 1$ if and only if edge $e \in E$. Similarly, let $\mathbf{y} \in \{0,1\}^{|\mathcal{P}(\tilde{E})|}$ determine paths in the graph $\mathcal{G}$, where $y_p = 1$ if and only if path $p \in \mathcal{P}(E)$. Lastly, for each pair $i, j \in V$, we define the error variables $\mathbf{z}_{ij} \in \{0,1\}^{|N_{ij}|}$, where $z_{ij}^{n} = 1$ if and only if the d-separation relation in $\mathcal{G}$ does *not* correspond to the independence relation found in the data (that is, $i \perp j | C_{ij}^{n}$ in $\mathcal{G}$ but $C_{ij}^{n} \in D_{ij}$, or $i \not\perp j | C_{ij}^{n}$ in $\mathcal{G}$ but $C_{ij}^{n} \in I_{ij}$).

[6] A DMG may technically contain paths with repeating nodes, due to cycles. However, as we show in Appendix B.1, all d-separation relations can be accurately captured by paths that conform to Definition 2. While this is a fairly technical result, it has important consequences for the correctness of our method, because it allows us to restrict attention to paths with non-repeating nodes without loss of inferential power.

We can now define the constraints and objective function of the model. First, we require two constraints that enforce coherence between the edge and path variables $\mathbf{x}$ and $\mathbf{y}$:

$$y_p \geq \sum_{e \in \tilde{E}} \phi_{pe} x_e - (\ell_p - 1), \quad p \in \mathcal{P}(\tilde{E}), \tag{3.4a}$$

$$\ell_p y_p \leq \sum_{e \in \tilde{E}} \phi_{pe} x_e, \quad p \in \mathcal{P}(\tilde{E}). \tag{3.4b}$$

The first constraint ensures that if all the edges on a path p are selected, then path p is present in the graph. The second constraint ensures that path p can only be present in the graph if all the edges on path p are selected. Next, note that a path p is blocked with respect to a set C_{ij}^n if and only if $\alpha_{ijp}^n = 0$. Thus, to satisfy all d-separation relations, we would ideally like to construct a graph such that for any $i, j \in V$, $\alpha_{ijp}^n y_p = 0$ holds for all $n \in N_{ij}^I$ and $p \in P_{ij}(\tilde{E})$. However, because the set of input d-separation and d-connection relations may not be jointly satisfiable, we allow for possible violations by introducing the error variable z_{ij}^n:

$$\alpha_{ijp}^n y_p \leq z_{ij}^n, \quad n \in N_{ij}^I,\ p \in P_{ij}(\tilde{E}),\ i, j \in V. \tag{3.5}$$

For each conditioning set $C_{ij}^n \in I_{ij}$, constraint (3.5) forces $z_{ij}^n = 1$ if i and j are *not* d-separated with respect to C_{ij}^n. Similarly, i and j are d-connected with respect to $C_{ij}^n \in D_{ij}$ if and only if there is at least one unblocked path, or equivalently, $\sum_{p \in P_{ij}(\tilde{E})} \alpha_{ijp}^n y_p \geq 1$. By again allowing for violations by introducing the error variable z_{ij}^n, we obtain the constraint

$$\sum_{p \in P_{ij}(\tilde{E})} \alpha_{ijp}^n y_p \geq 1 - z_{ij}^n, \quad n \in N_{ij}^D,\ i, j \in V, \tag{3.6}$$

which forces $z_{ij}^n = 1$ if i and j are not d-connected with respect to $C_{ij}^n \in D_{ij}$. Note that constraint (3.5) and (3.6) are defined over N_{ij}^I and N_{ij}^D, respectively, so that each z_{ij}^n variable appears in one

constraint only. It follows from (3.5) and (3.6) that the total number of violated d-connection and d-separation relations is given by $\sum_{i,j \in V} \sum_{n \in N_{ij}} z_{ij}^n$. Minimizing these violations over the constraints (3.4)–(3.6) yields the optimization problem:

$$
\begin{aligned}
& \underset{\mathbf{x},\mathbf{y},\mathbf{z}}{\text{minimize}} && \sum_{i,j \in V} \sum_{n \in N_{ij}} z_{ij}^n \\
& \text{subject to} && (3.4) - (3.6), \\
\text{CausalIP}(\tilde{E}): & && y_p \in \{0,1\}, \quad p \in \mathcal{P}(\tilde{E}), \\
& && z_{ij}^n \in \{0,1\}, \quad n \in N_{ij},\ i,j \in V, \\
& && x_e \in \{0,1\}, \quad e \in \tilde{E}.
\end{aligned}
$$

We refer to the formulation above as CausalIP($\tilde{E}$), where $\tilde{E}$ is the set of edges the model has access to. If $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ is a solution to CausalIP($\tilde{E}$), then the graph returned by the model is given $\mathcal{G}(\mathbf{x}) = (V, E)$ where $E = \{e | e \in \tilde{E} \text{ and } x_e = 1\}$. In Appendix B.2, we show how our approach can also be restricted to DAG- or ADMG-search (although more scalable methods exist for those spaces). Next, we address the theoretical performance of CausalIP($\tilde{E}$).

3.3.2 Discovery Guarantee

In general, the independence structure seen in observational data does not uniquely identify the underlying causal graph. Two graphs that have the same independence structure (and thus the same d-separation relations) are said to be *Markov equivalent*.

Definition 5 (Markov equivalence)**.** *DMG $\mathcal{G}_1 = (V, E)$ is Markov equivalent to DMG $\mathcal{G}_2 = (V, E')$ if and only if $\mathcal{G}_1$ and $\mathcal{G}_2$ have identical d-separation relations:*

$$
\mathcal{G}_1 \sim \mathcal{G}_2 \quad \text{if and only if} \quad i \perp_{\mathcal{G}_1} j | C \iff i \perp_{\mathcal{G}_2} j | C \quad \forall i, j \in V \text{ and } C \subseteq V \setminus \{i, j\}. \tag{3.8}
$$

Two graphs that are Markov equivalent cannot be distinguished by their d-separation relations alone.[7] In the context of causal discovery, the Markov equivalence class of the true, data-generating graph is the limit of what can be learned about the causal structure from the independence structure in the data (Geiger and Pearl, 1988; Meek, 1995).[8]

We now show that, under appropriate conditions, CausalIP correctly uncovers a graph in the Markov equivalence class of the true causal graph. We assume that we have access to the results of all possible independence tests over a given set of variables, and that the test results correctly describe an underlying ground truth DMG $\mathcal{G}_T$:

Assumption 6 (Complete oracle)**.** *Let $\mathcal{G}_T$ be the true data-generating graph. For all $i, j \in V$ and $C \subseteq V \setminus \{i, j\}$: (i) $C \in D_{ij}$ if and only if $i \not\perp\!\!\!\perp j | C$ in $P_{\mathcal{G}_T}(V)$, and (ii) $C \in I_{ij}$ if and only if $i \perp\!\!\!\perp j | C$ in $P_{\mathcal{G}_T}(V)$.*

Assumption 6 allows us to separate the discovery task, handled by CausalIP, from the statistical inference of the conditional independence tests. The assumption also describes the model inputs (i.e., I_{ij} and D_{ij}) that would be obtained in the large-sample limit, which we use to prove the asymptotic correctness of our model. Next, let

$$E^c = \{i \leftarrow j,\ i \rightarrow j,\ i \leftrightarrow j,\ \forall i, j \in V : i \neq j\} \tag{3.9}$$

be the set of all possible directed and bi-directed edges in a complete graph over V. We can now state the main result of this section:

Proposition 3. *Let $\mathcal{G}^c$ be the graph returned by* CausalIP*(E^c) given Assumption 6. Then* (i)

[7] In the acyclic, causally sufficient case (i.e., for DAGs), the features shared by Markov equivalent DAGs can be easily characterized: two DAGs are Markov equivalent if and only if they share the same *skeleton* (unoriented adjacency structure) and the same *unshielded colliders* (Verma and Pearl (1991)). For the general class of DMGs that we are considering here, no such compact characterization of Markov equivalent graphs exists.

[8] To further distinguish Markov equivalent graphs requires experimental intervention, stronger background assumptions, or that one can make assumptions about the data distribution that go beyond its independence structure (e.g., about particular parameterizations; see e.g. Eberhardt (2017) for an overview).

$\mathcal{G}^c \in \mathcal{O}$ (i.e., $\mathcal{G}^c$ minimizes the objective in (3.3)) and (ii) $\mathcal{G}^c \sim \mathcal{G}_T$ (i.e., $\mathcal{G}^c$ and $\mathcal{G}_T$ are Markov equivalent).

Proposition 3 confirms the asymptotic correctness of CausalIP (all proofs are contained in Appendix B.4). When V is small, it is possible to solve CausalIP over the complete set of edges E^c using off-the-shelf integer programming solvers. However, this approach is not viable even for moderately sized graphs ($|V| > 6$), because CausalIP(E^c) searches over *all* possible paths induced by the set E^c, which explodes combinatorially in $|V|$. For larger graphs, a more efficient way of selecting paths is required.

3.4 Edge Generation Algorithm

Rather than attempt to solve CausalIP(E^c) over the complete set of edges E^c, we build an iterative procedure that efficiently constructs a set of *candidate edges* $\tilde{E} \subset E^c$. The selection of new edges is based on the observation that every path p (with $\ell_p \geq 2$) can be represented as a concatenation of node triples (i, j, k) that either form a collider at j or a non-collider at j. We refer to such a triple as a *collider chain* or a *non-collider chain*, respectively (see Figure 3.3). Next, we present necessary conditions for the presence of collider and non-collider chains in the underlying causal graph:

Lemma 1 (Necessary conditions for chains)**.** *Suppose Assumption 6 holds and consider a triple of nodes (i, j, k) in a graph $\mathcal{G} \sim \mathcal{G}_T$. Then*

(i) If there exists a collider chain over (i, j, k) in $\mathcal{G}$, then $I_{ij} = I_{jk} = \emptyset$ and $j \notin C$ for all $C \in I_{ik}$.

(ii) If there exists a non-collider chain over (i, j, k) in $\mathcal{G}$, then $I_{ij} = I_{jk} = \emptyset$ and $j \in C$ for all $C \in I_{ik}$.

Lemma 1 formalizes the intuition that every triple along every path of the underlying causal graph will leave its corresponding signature in the independence structure of the data. These

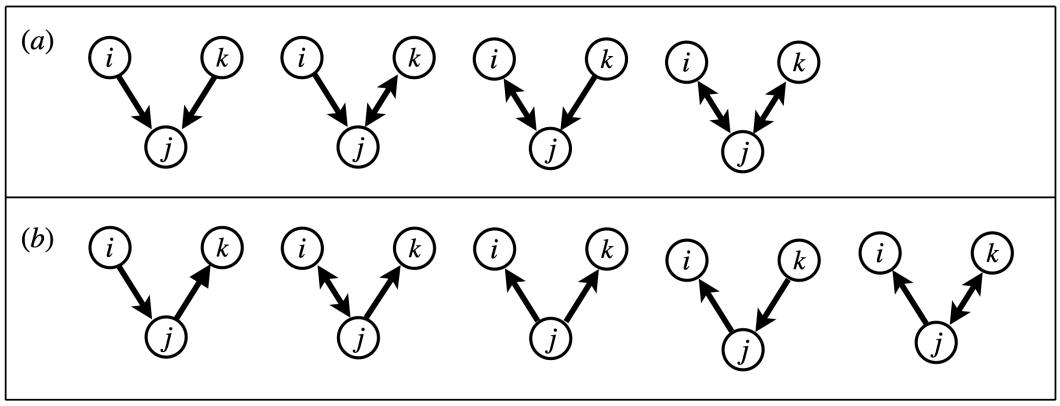


Figure 3.3: Collider chains (a) and non-collider chains (b).

signatures serve as indicators of the potential presence of the corresponding chains in the graph. The conditions in Lemma 1 do not *guarantee* that the underlying graph contains the corresponding chain, because they can also be generated by other causal structures (see Figure 3.4). In other words, if a triple (i, j, k) satisfies the conditions in Lemma 1(i) or (ii), we interpret this as strong (but not conclusive) evidence of the corresponding chain's presence in the underlying causal graph. Using this characterization of chains, we construct the following two sets:

$$S = \{(i, j, k) | I_{ij} = I_{jk} = \emptyset \text{ and } j \notin C \text{ for all } C \in I_{ik}\}, \tag{3.10}$$

$$\bar{S} = \{(i, j, k) | I_{ij} = I_{jk} = \emptyset \text{ and } j \in C \text{ for all } C \in I_{ik}\}. \tag{3.11}$$

Here, S and $\bar{S}$ contain all triples (i, j, k) whose independence structure is indicative of a collider and non-collider chain, respectively. These sets are not disjoint in general: A triple (i, j, k) where (i, k) are never conditionally independent (i.e., $I_{ik} = \emptyset$) will be included in both S and $\bar{S}$.

The sets S and $\bar{S}$ contain triples that plausibly form chains in the underlying graph, and thus indicate plausible edges as well. We therefore focus on edges represented by the sets S and $\bar{S}$ when constructing the set of candidate edges $\tilde{E}$.

We further refine our search for edges by also considering d-connection or d-separation relations

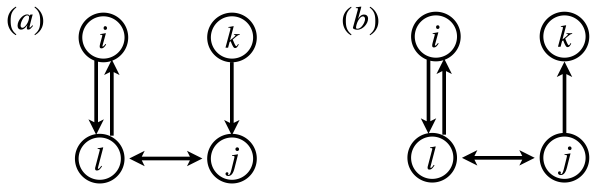


Figure 3.4: **(a)** The triple (i, j, k) satisfies all conditions in Lemma 1(i) but does not form a collider chain. **(b)** The triple (i, j, k) satisfies all conditions in Lemma 1(ii) but does not form a non-collider chain.

that are unsatisfied by an incumbent solution. For an initial set of candidate edges $\tilde{E}$, let $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ be the solution to CausalIP($\tilde{E}$) with the corresponding graph $\mathcal{G}(\mathbf{x})$. The error variable $\mathbf{z}$ tracks those d-separations and d-connections that are inconsistent with the independence and dependence findings in the data. Since our method will incrementally add edges to $\tilde{E}$, we focus on those errors where a conditional dependence found in the data is not yet matched by a d-connection in the graph $\mathcal{G}(\mathbf{x})$. Specifically, we are interested in the pairs (i, k) where we have $C_{ik}^n \in D_{ik}$ based on the test results, but the d-connection $i \not\perp k | C_{ik}^n$ is *not* satisfied in $\mathcal{G}(\mathbf{x})$, and consequently $z_{ik}^n = 1$. Accordingly, for each pair (i, k), we can define the set

$$N_{ik}^D(\mathbf{z}) = \{n \in N_{ik}^D | z_{ik}^n = 1\} \tag{3.12}$$

to represent the d-connection relations for (i, k) that are implied by the conditional independence tests but violated by the current graph $\mathcal{G}(\mathbf{x})$. Since our goal is to select *new* edges to add to $\tilde{E}$, we have to identify edges that are not already in $\tilde{E}$. To that end, let E_{ijk} and $\bar{E}_{ijk}$ be the sets of all possible edges in a chain over (i, j, k), in which j is a collider or non-collider, respectively:

$$E_{ijk} = \big\{i \to j,\ i \leftrightarrow j,\ j \leftarrow k,\ j \leftrightarrow k\big\}, \tag{3.13a}$$

$$\bar{E}_{ijk} = \big\{i \leftarrow j,\ i \to j,\ i \leftrightarrow j,\ j \leftarrow k,\ j \to k,\ j \leftrightarrow k\big\} \tag{3.13b}$$

Here, E_{ijk} is the set of edges that define the collider chains in Figure 3.3(a), and $\bar{E}_{ijk}$ is the set of

edges that define the non-collider chains in Figure 3.3(b). We can now identify triples that satisfy two criteria: (i) some of their edges have not yet been considered (i.e., $E_{ijk} \not\subset \tilde{E}$ or $\bar{E}_{ijk} \not\subset \tilde{E}$), and (ii) in the incumbent graph $\mathcal{G}(\mathbf{x})$, i and k are d-separated for some conditioning set C_{ik}^n ($i \perp k | C_{ik}^n$), even though test results indicate that i and k are dependent for that conditioning set ($i \not\perp k | C_{ik}^n$). We now define two sets that contain triples that satisfy these two criteria:

$$\Psi(\mathbf{z}) = \{(i,j,k) \mid \text{there exists } n \in N_{ik}^D(\mathbf{z}) \text{ such that } j \in C_{ik}^n \text{ and } E_{ijk} \not\subset \tilde{E}\}, \tag{3.14a}$$

$$\bar{\Psi}(\mathbf{z}) = \{(i,j,k) \mid \text{there exists } n \in N_{ik}^D(\mathbf{z}) \text{ such that } j \notin C_{ik}^n \text{ and } \bar{E}_{ijk} \not\subset \tilde{E}\}. \tag{3.14b}$$

The sets $\Psi(\mathbf{z})$ and $\bar{\Psi}(\mathbf{z})$ need not be disjoint. They only differ in their check of whether j belongs to a conditioning set C_{ik}^n that shows i and k to be dependent in the data. This specific check on the role of j ensures that if we now combine the Ψ-sets with the S-sets, the collider/non-collider chains associated with the triples of the respective sets are candidates to address the inconsistencies identified by $\mathbf{z}$. We can finally define the sets that are the focal points of our search for new edges:

$$S(\mathbf{z}) = S \cap \Psi(\mathbf{z}), \tag{3.15a}$$

$$\bar{S}(\mathbf{z}) = \bar{S} \cap \bar{\Psi}(\mathbf{z}), \tag{3.15b}$$

Intuitively, $S(\mathbf{z})$ and $\bar{S}(\mathbf{z})$ represent chains that we have good reason to think exist in the graph (because of S and $\bar{S}$) *and* correspond to edges useful for satisfying a required d-connection that is violated by $\mathcal{G}(\mathbf{x})$ (because of $\Psi(\mathbf{z})$ and $\bar{\Psi}(\mathbf{z})$).

We have thus far established that edges among the triples in $S(\mathbf{z})$ and $\bar{S}(\mathbf{z})$ are strong candidates for inclusion in the candidate set $\tilde{E}$ to send to CausalIP($\tilde{E}$). We now address the problem of how to actually select edges from these sets to pass to CausalIP($\tilde{E}$).

3.4.1 Generating Candidate Edges

Since the tractability of CausalIP($\tilde{E}$) suffers if the edge set $\tilde{E}$ is too large, we would ideally like to select edges that can reconcile unsatisfied d-connections, without needlessly introducing redundant edges. To that end, our approach will be to select the smallest number of edges such that at least one edge from each triple in $S(\mathbf{z})$ and $\bar{S}(\mathbf{z})$ is selected. This minimal edge selection problem is itself a combinatorial optimization problem, because the chains in $S(\mathbf{z})$ and $\bar{S}(\mathbf{z})$ may share common edges. Accordingly, we formulate it as an integer program, which we call NewEdgesIP.

There are three types of edges that may exist between every pair of nodes (i, j): $i \to j, i \leftarrow j$ and $i \leftrightarrow j$. We index these three edge types by $t \in \{1, 2, 3\}$, respectively. Let w_{ij}^t be a binary decision variable where $w_{ij}^t = 1$ if a type t edge between nodes i and j is selected to be included in $\tilde{E}$, and $w_{ij}^t = 0$ otherwise. Let λ_{ij}^t be a parameter where $\lambda_{ij}^t = 1$ if $\tilde{E}$ contains a type t edge between nodes i and j, and $\lambda_{ij}^t = 0$ otherwise.

We now define the constraints and objective of NewEdgesIP. To select edges from $S(\mathbf{z})$, we include the following constraints:

$$\sum_{t \in \{1,3\}} (w_{ij}^t + \lambda_{ij}^t) \geq 1, \quad (i, j, k) \in S(\mathbf{z}), \tag{3.16a}$$

$$\sum_{t \in \{2,3\}} (w_{jk}^t + \lambda_{jk}^t) \geq 1, \quad (i, j, k) \in S(\mathbf{z}), \tag{3.16b}$$

$$\sum_{t \in \{1,3\}} w_{ij}^t + \sum_{t \in \{2,3\}} w_{jk}^t \geq 1, \quad (i, j, k) \in S(\mathbf{z}). \tag{3.16c}$$

The first two constraints construct a collider chain over (i, j, k): the first ensures that there is either a new or existing edge between i and j with an arrowhead at j, and the second ensures there is either a new or existing edge between k and j with an arrowhead at j. Then, the third constraint forces at least one new edge to be selected from the candidate collider chain. Similarly, to select

edges from $\bar{S}(\mathbf{z})$, we include:

$$\sum_{t\in\{1,2,3\}} (w_{ij}^t + \lambda_{ij}^t) \geq 1, \quad (i,j,k) \in \bar{S}(\mathbf{z}), \tag{3.17a}$$

$$\sum_{t\in\{1,2,3\}} (w_{jk}^t + \lambda_{jk}^t) \geq 1, \quad (i,j,k) \in \bar{S}(\mathbf{z}), \tag{3.17b}$$

$$w_{ij}^2 + \lambda_{ij}^2 + w_{jk}^1 + \lambda_{jk}^1 \geq 1, \quad (i,j,k) \in \bar{S}(\mathbf{z}), \tag{3.17c}$$

$$\sum_{t\in\{1,2,3\}} (w_{ij}^t + w_{jk}^t) \geq 1, \quad (i,j,k) \in \bar{S}(\mathbf{z}). \tag{3.17d}$$

Analogous to (3.16a) and (3.16b), the first three constraints above construct a non-collider chain over (i,j,k), using either existing or new edges: The first two constraints ensure an edge exists between both (i,j) and (j,k), and the third constraint ensures that j is a non-collider. Then, the fourth constraint forces at least one new edge to be selected from the constructed non-collider chain.

The final group of constraints we include are:

$$w_{ij}^t + \lambda_{ij}^t \leq 1, \quad i,j \in V,\ t \in \{1,2,3\}, \tag{3.18a}$$

$$w_{ij}^1 = w_{ji}^2, \quad i,j \in V, \tag{3.18b}$$

$$w_{ij}^3 = w_{ji}^3, \quad i,j \in V. \tag{3.18c}$$

The first constraint ensures we do not select an edge that is already included in $\tilde{E}$. The second constraint enforces that $i \to j$ and $j \leftarrow i$ are the same edge, and the third constraint enforces that $i \leftrightarrow j$ and $j \leftrightarrow i$ are the same edge.

Our objective is to minimize the total number of new edges added to the set of candidate edges $\tilde{E}$. Combining this objective with the constraints (3.16)–(3.18) and forcing each w_{ij}^t to be binary yields the following formulation:

$$\text{minimize}_{\mathbf{w}} \quad \sum_{i,j \in V} \sum_{t \in \{1,2,3\}} w_{ij}^t$$

NewEdgesIP$(S(\mathbf{z}), \bar{S}(\mathbf{z}))$: subject to (3.16) − (3.18)

$$w_{ij}^t \in \{0,1\}, \quad i,j \in V,\ t \in \{1,2,3\}.$$

The output of this formulation is a set of new edges E_{new} to be added to $\tilde{E}$, where E_{new} contains a type t edge between nodes i and j if and only if $w_{ij}^t = 1$ at an optimal solution to NewEdgesIP. In summary, NewEdgesIP generates edges efficiently by searching for edges that satisfy the following criteria:

(i) the edges belong to collider or non-collider chains for which we have strong evidence of their presence in the true graph based on the observed independence and dependence relations (i.e., the chains belong to S or $\bar{S}$), and

(ii) The edges belong to collider or non-collider chains whose inclusion in the graph would satisfy a d-connection relation violated by the incumbent solution (i.e., the chains belong to $\Psi(\mathbf{z})$ or $\bar{\Psi}(\mathbf{z})$).

Having defined the key components of our method, we now present a summary of the algorithm and prove its correctness.

3.4.2 Algorithm Summary and Main Result

Algorithm 3 provides an overview of the main steps. In the main loop, the algorithm iterates between calling the sub-algorithm UpdateEdges to generate new edges to add to $\tilde{E}$, and solving CausalIP$(\tilde{E})$ to identify unsatisfied d-separation and d-connection relations. The algorithm terminates and returns the graph $\mathcal{G}^*$ when all d-connection and d-separation relations are satisfied, or a generic alternate termination criterion (represented by Term in Algorithm 3) is satisfied.

Algorithm 4 describes the UpdateEdges sub-algorithm that is used to generate new edges, which primarily involves solving NewEdgesIP. In the case where $S(\mathbf{z}) = \bar{S}(\mathbf{z}) = \emptyset$, UpdateEdges randomly picks triples from S and $\bar{S}$ to pass to NewEdgesIP. To capture edges that may not belong to any chain in the true graph (i.e., those that do not correspond to any member of S or $\bar{S}$), we initialize the set of candidate edges as $\tilde{E}_0$:

$$\tilde{E}_0 = \{i \to j, i \leftarrow j, i \leftrightarrow j \mid I_{ij} = \emptyset, D_{ik} = D_{jk} = \emptyset \text{ for all } k \in V \setminus \{i, j\}\}.$$

Algorithm 3: EdgeGen.

Input: V, S, $\bar{S}$.
Output: $\mathcal{G}^*$.
Initialize: $\tilde{E} = \tilde{E}_0$,
1. Solve CausalIP($\tilde{E}$) to obtain solution $(\mathbf{x}, \mathbf{y}, \mathbf{z})$.
2. **while** $\sum_{i,j \in V} \sum_{n \in N_{ij}} z_{ij}^n > 0$ **and** Term=false:
   Call UpdateEdges to update $\tilde{E}$.
   Solve CausalIP($\tilde{E}$) to obtain solution $(\mathbf{x}, \mathbf{y}, \mathbf{z})$.
3. Set $E^* = \{e \in \tilde{E} | x_e = 1\}$ and return $\mathcal{G}^* = (V, E^*)$.

Algorithm 4: UpdateEdges sub-algorithm.

Input: $S(\mathbf{z}), \bar{S}(\mathbf{z}), S, \bar{S}, \tilde{E}$.
Output: Updated candidate edges $\tilde{E}$.
1. **if** $S(\mathbf{z}) \cup \bar{S}(\mathbf{z}) \neq \emptyset$:
   Solve NewEdgesIP($S(\mathbf{z}), \bar{S}(\mathbf{z})$) to obtain E_{new}.
   else:
   Pick any $(i, j, k) \in S \cap \bar{S}$ such that $E_{ijk} \not\subset \tilde{E}$. Set $R(\mathbf{z}) = (i, j, k)$.
   Pick any $(i, j, k) \in S \cap \bar{S}$ such that $\bar{E}_{ijk} \not\subset \tilde{E}$. Set $\bar{R}(\mathbf{z}) = (i, j, k)$.
   if $R(\mathbf{z}) = \emptyset$ and $\bar{R}(\mathbf{z}) = \emptyset$:
   Pick any $(i, j, k) \in S$ such that $E_{ijk} \not\subset \tilde{E}$. Set $R(\mathbf{z}) = (i, j, k)$.
   Pick any $(i, j, k) \in \bar{S}$ such that $\bar{E}_{ijk} \not\subset \tilde{E}$. Set $\bar{R}(\mathbf{z}) = (i, j, k)$.
   Solve NewEdgesIP($R(\mathbf{z}), \bar{R}(\mathbf{z})$) to obtain E_{new}.
2. Update $\tilde{E} \leftarrow \tilde{E} \cup E_{\text{new}}$.

In the setting where there exists a directed mixed graph that can satisfy all input constraints, EdgeGen is guaranteed to terminate with such a graph.

Proposition 4. *Under Assumption 6,* EdgeGen *terminates with* $\sum_{i,j \in V} \sum_{n \in N_{ij}} z_{ij}^n = 0$.

When the input relations are not jointly satisfiable, we require an alternate termination condition. A straightforward approach, which we use in our numerical experiments, is to terminate EdgeGen after a fixed number of iterations with no improvement in the number of unsatisfied input constraints (i.e., the optimal objective of CausalIP($\tilde{E}$)). We can now formally state our main theoretical result, which builds on Proposition 4:

Theorem 5. *Let* $\mathcal{G}^*$ *be the graph returned by* EdgeGen *given Assumption 6. Then* (*i*) $\mathcal{G}^* \in \mathcal{O}$ *(i.e.,* $\mathcal{G}^*$ *minimizes the objective in (3.3) and* (*ii*) $\mathcal{G}^* \sim \mathcal{G}_T$ *where* $\mathcal{G}_T$ *is true underlying graph (i.e.,* $\mathcal{G}^*$ *and* $\mathcal{G}_T$ *are Markov equivalent).*

Theorem 5 states that EdgeGen, which is far more scalable than a brute-force solution of CausalIP(E^c), preserves the same discovery guarantees given in Proposition 3.

Similar to Proposition 3, Theorem 5 is an asymptotic guarantee, due to its dependence on Assumption 6. In the more realistic finite-sample setting where Assumption 6 does not hold, the input relations may not be jointly satisfiable. In this setting, most constraint-based methods that also allow for cycles and confounders aim to find a graph that minimizes (weighted) violations of the input constraints (Hyttinen et al., 2013, 2014, 2017; Rantanen et al., 2020). However, *exactly* minimizing such violations requires searching over the entire space of DMGs, which, in the pursuit of improved scalability, we deliberately avoid (by only considering a subset of possible edges $\tilde{E}$ instead of the complete edge set E^c). As a consequence, in the finite-sample setting where Assumption 3 does not hold, our approach can be viewed as a heuristic for minimizing the number of unsatisfied dependence and independence relations. As demonstrated in the numerical results below, the advantage of this heuristic approach is that it scales to instances that are intractable for provably optimal methods from the literature, while maintaining reasonable accuracy. Further, in the setting

where Assumption 6 *does* hold, our method outperforms appropriate benchmark algorithms by an order of magnitude with respect to solution time, without sacrificing optimality.

3.4.3 Computational Performance

In this section we examine the computational performance of EdgeGen using synthetic data. To serve as performance benchmarks, we also implemented the causal discovery methods described in Hyttinen et al. (2013) and Hyttinen et al. (2014), both of which also allow for feedback loops and latent confounders.[9] Hyttinen et al. (2013) solve the discovery problem using a Boolean satisfiability solver, and Hyttinen et al. (2014) propose a solution method based on answer set programming; for conciseness, we will refer to these two approaches as SAT and ASP, respectively. We also created an additional benchmark by combining the logical encoding developed in Hyttinen et al. (2013) and the solver used in Hyttinen et al. (2014), which we refer to as SAT+ASP.

Setup.

We conducted two sets of numerical experiments. First, we considered a *conflict-free* setting in which there are no conflicts among the input d-separation and d-connection relations (i.e., Assumption 6 holds). Because all four methods are guaranteed to return a graph that is Markov equivalent to the ground truth graph $\mathcal{G}_T$ in this setting, we focus our comparison exclusively on solution times. Second, we considered a *conflicted* setting where the input constraints are not jointly satisfiable (i.e., Assumption 6 does not hold). In this setting, the SAT algorithm does not apply, so we compare EdgeGen with ASP with respect to solution time and accuracy.

For the conflict-free case, we generated 50 random directed graphs for $|V| = 5, 6, \ldots, 15$. These graphs were generated to have an average degree of 3, and were permitted to have cycles and

[9] Hyttinen et al. (2017) and Rantanen et al. (2020) are also relevant benchmarks here, but we do not compare against them because implementable code is not publicly available for those methods.

unobserved confounders. For each graph, we computed all d-separation and d-connection relations, which are the required inputs for all four methods. Note that in the conflict-free setting, the data generation procedure is irrelevant, because we have direct knowledge of the d-separation and d-connection relations of the ground-truth graphs.

For the conflicted case, we set up the experiment using the code from Hyttinen et al. (2014). We generated 50 linear Gaussian models with $|V| = 4, 5, \ldots, 9$. In this setup, the independence and dependence relations for each model are computed using correlation based t-tests over 500 samples, with a significance level of 0.01. We ran all experiments on an Intel Xeon E5-2680 machine with 3.0GHz×24 processors and 20 GB of memory, and used Gurobi v8.0 to solve CausalIP and NewEdgesIP. We terminated EdgeGen if no improvement in the objective was observed after three consecutive iterations. Because all methods require the same pre-computation of d-separation and d-connection relations (either from a ground truth graph or through conditional independence tests), we isolate the performance of each of the four methods by reporting the solution times of the discovery task only.

Results.

Table 3.1 summarizes the median solution times over 50 instances for EdgeGen, SAT, ASP and SAT+ASP, and demonstrates how the solution times increase in $|V|$ for all four methods. The impact of the instance size on solution time is most pronounced for SAT and ASP, which are unable to scale past $|V| = 11$ and $|V| = 13$ nodes, respectively, consistent with the results reported in those papers. The SAT+ASP procedure is superior to both of its constituent methods with respect to solution time, although it cannot scale past $|V| = 14$ variables, also due to insufficient memory. By comparison, EdgeGen significantly reduces solution times (and memory usage), allowing it to solve instances with $|V| = 15$ nodes in 30 seconds on average. We note here that a brute-force

solution of CausalIP(E^c) is unable to scale past $|V| = 6$ nodes (results not shown). This suggests that the efficiency of EdgeGen is due to the edge generation procedure, and not a consequence of formulating the discovery problem as an integer program.

An important observation in Table 3.1 is that the main computational bottleneck in SAT, ASP and SAT+ASP is the discovery task itself. For these three benchmark methods, one could argue that the time required to compute all d-separation and d-connection relations is irrelevant, because they encounter memory issues during the discovery phase for larger instances. In contrast, Table 3.1 shows that EdgeGen scales gracefully with respect to solution time on the discovery task. We did not test instances larger than $|V| = 15$ because the time required to compute the (exponential number of) d-separation and d-connection relations becomes a bottleneck for all four methods (e.g., for $|V| = 16$, they cannot be computed in under two hours).

$\|V\|$	EdgeGen	SAT	ASP	SAT+ASP
5	0	0	0	0
6	0	1	0	0
7	0	4	0	0
8	0	15	2	1
9	0	64	6	2
10	1	158	21	6
11	1	615	61	16
12	3	-	197	36
13	8	-	619	98
14	27	-	-	260
15	101	-	-	-

Table 3.1: Median solution times (nearest CPU second) over 50 random instances in conflict-free setting. Dashes indicate instance could not solve due to insufficient memory (20GB).

Table 3.2 summarizes the results for the case with conflicted independence relations. Because SAT and SAT+ASP are not designed to handle conflicts, we report the median solution time and accuracy of EdgeGen and ASP only. We report the accuracy of each algorithm using two performance metrics. *Loss* refers to the fraction of independence relations that are unsatisfied by

the returned graph (for EdgeGen, *Loss* is the value of the CausalIP objective at termination.) *Error* refers to the fraction of d-separation and d-connection relations from the *true* graph that are unsatisfied. This distinction arises because the independence relations obtained from data may not reflect the d-separation relations of the true graph due to errors in statistical testing.

Table 3.2 shows that (the provably optimal) ASP algorithm outperforms EdgeGen with respect to accuracy for small graphs, but cannot scale beyond $|V| = 6$. The EdgeGen algorithm scales up to $|V| = 9$, and reaches the termination criterion in under 1 minute for a majority of instances, although the solutions are not optimal. For $|V| = 10$, EdgeGen frequently reached the 1 hour time limit, so we did not conduct comprehensive tests for graphs of that size. Note that ASP outperforms EdgeGen with respect to both accuracy measures for small graphs. This gap related to tractability is due to our choice to to terminate EdgeGen before exhausting all possible edges, which restricts the search space of possible graphs. More comprehensive experiments may reveal the extent to which each of these factors contribute to the accuracy gap between EdgeGen and ASP. Regardless, for larger instances ($|V| \geq 6$), the iterative approach of EdgeGen allows it to produce graphs with reasonable accuracy for problem sizes where ASP is unable to return any graph at all.

In summary, Tables 3.1 and 3.2 suggest that EdgeGen offers substantial advantages over the benchmarks with respect to both solution time and memory usage. The efficiency of our approach is most apparent in the conflict-free setting (Table 3.1), where all methods produce optimal solutions, but EdgeGen dramatically outperforms all three benchmarks. In the conflicted setting (Table 3.2), the trade-off is clear – EdgeGen is less computationally demanding with respect to memory, allowing it to handle larger instances, but unlike ASP, the output graphs are not provably optimal.

	EdgeGen			ASP		
$\|V\|$	Time	Loss	Error	Time	Loss	Error
5	0	0.12	0.17	0.08	0.08	0.12
6	0	0.16	0.18	37.05	0.11	0.12
7	1	0.17	0.23	-	-	-
8	13	0.21	0.26	-	-	-
9	31	0.22	0.27	-	-	-
10	87	0.27	0.33	-	-	-

Table 3.2: Median solution times (rounded to nearest CPU second), loss, and error over 50 random instances in conflicted setting. Dashes indicate instance could not solve due to insufficient memory (20GB).

3.5 Feature Selection Using Markov Blankets

Thus far, our focus has been on constructing a complete causal graph over a set of observed variables, in which all causal relations are deemed equally important. In empirical applications, however, there may be a *target variable* whose causes are of particular interest to the researcher (e.g., income or health outcomes). An intuitive approach to investigating causality with respect to a target variable is to simply construct the graph over all observed variables, which then reveals causal pathways related to the target variable. However, if the number of variables is large, this naive approach may be computationally inefficient or intractable. Instead, if we can *a priori* identify a subset of variables that collectively carry all useful information about the target variable, then we can sidestep the computational burden of learning causal relations among variables of secondary importance. The concept of a *Markov blanket* is a formalization of this idea:

Definition 6 (Markov blankets)**.** *The Markov blanket of a target variable T, $MB(T)$, is a minimal set of variables in $V \setminus \{T\}$ such that*

$$i \perp\!\!\!\perp T | MB(T) \text{ for all } i \in V \setminus \{MB(T), T\}. \tag{3.20}$$

Given a target variable T, a Markov blanket $MB(T)$ is the smallest set of variables such that all

other variables are probabilistically independent of T conditional on $MB(T)$. In machine learning contexts where T is the variable to be predicted (e.g., a class label), the Markov blanket of T is the smallest subset of features that has the same predictivity as the full set of features. Accordingly, identifying a Markov blanket based on causal structures is often referred to as *causal feature selection* (Aliferis et al., 2010). Further, under the usual correspondence between independence and d-separation (Remark 1), the Markov blanket of a target variable T contains the direct causes and effects of T (in addition to other nodes). To that end, learning Markov blankets can be viewed as a step toward discovering the *local* causal structures around a target variable T, which is the perspective we adopt here.

In this section, we present an optimization model for returning a Markov blanket of an input target variable T based purely on the independence structure of the data. This approach can then be used as a pre-processing step to be run before the EdgeGen algorithm, with the aim of eliminating potentially redundant variables from the discovery task before searching for the local DMG structure around T.

3.5.1 Related Literature

There are two general approaches to finding Markov blankets. Aliferis et al. (2010) denote them as *causal* vs. *non-causal* methods, to distinguish whether the search for the Markov blanket requires the identification of the (local) causal structure of the target variable prior to identification of the Markov blanket, or not. Since our motivation is to use the Markov blanket as an aid to focus the local causal discovery, our approach to searching for the Markov blanket itself has to be non-causal. This approach also sidesteps one of the challenges of extant *causal* methods to find the Markov blanket that have to make assumptions about the nature of the underlying causal structure. Most of these methods assume acyclicity and causal sufficiency, making them inapplicable to the more

general setting we are considering in this work.

In the causally sufficient setting, the first provably correct algorithm in discovering the Markov blanket of a target variable is introduced by Margaritis and Thrun (2000). Several other methods have aimed to improve the efficiency and scalability of Markov blanket discovery under various assumptions on the underlying class of graphs and available data (Tsamardinos et al. (2003), Yaramakala and Margaritis (2005), Tsamardinos et al. (2006), Pena et al. (2007), Aliferis et al. (2010), Gao and Ji (2015), Wu et al. (2019), Yu et al. (2019)).

Previous work on Markov blankets in causally insufficient settings is limited. Richardson (2003) and Pellet and Elisseeff (2008) provide characterizations of Markov blankets for classes of graphs that permit unobserved confounders. Two recently proposed approaches for learning Markov blankets in the presence of confounding are by Yu et al. (2018) and Triantafillou et al. (2021), although both focus on acyclic graphs.

3.5.2 Learning Markov Blankets with Optimization

Given that we assume the results of all possible conditional independence tests are available, identifying the Markov blanket is equivalent to searching for the minimal conditioning set that satisfies (3.20), which can be done by simply sorting all possible conditional sets. However, we formulate the search for a Markov blanket as an optimization problem, which allows us to easily include additional considerations, namely, balancing the number of selected variables with the number of violations of the Markov blanket condition (3.20).

Let m_i be a binary decision variable where $m_i = 1$ if and only if node i is selected to be included in the Markov blanket. For convenience, we will write $V(T) = \{i \in V | m_i = 1\}$ to denote the set of nodes selected by the model. For each $i \in V$ and $n \in N_{Ti}^I$, let v_i^n be a binary decision variable where $v_i^n = 1$ if and only if $V(T)$ is equal to the n^th conditioning set in A_{Ti}, i.e. $V(T) = C_{Ti}^n$, and

define a parameter θ_{ij}^n where $\theta_{ij}^n = 1$ if and only if $j \in C_{Ti}^n$. We define the following two constraints to enforce the correspondence between $\mathbf{m}$ and $\mathbf{v}$:

$$v_i^n \leq m_j, \quad j \in C_{Ti}^n,\ n \in N_{Ti}^I,\ i \in V \setminus \{T\}, \tag{3.21a}$$

$$m_j + v_i^n \leq \theta_{ij}^n + 1, \quad n \in N_{Ti}^I,\ i, j \in V \setminus \{T\}. \tag{3.21b}$$

The first constraint ensures that if $v_i^n = 1$ (equivalently, if $V(T) = C_{Ti}^n$), then all nodes $j \in C_{Ti}^n$ are selected. The second constraint ensures that if $V(T) = C_{Ti}^n$, then the nodes not in C_{Ti}^n cannot be in $V(T)$. Note that constraint (3.21b) is active only when $v_i^n = 1$ and $\theta_{ij}^n = 0$, which forces m_j to be zero. Definition 6 for the Markov blanket is then satisfied if we minimize the number of selected variables – given by $\sum_{i \in V \setminus \{T\}} m_i$ – subject to (3.21) and the following constraint:

$$m_i + \sum_{n \in N_{Ti}^I} v_i^n \geq 1. \tag{3.22}$$

This would simply implement the search over conditioning sets to find the exact Markov blanket. However, doing so would neglect the practical challenges a user often faces, such as trading-off between the number of variables selected, and their predictivity with respect to the target T compared to the full set of variables V. In particular, while Markov blankets in a DAG consist just of the *parents, children* and *spouses* (other parents of the children) of the target variable T (Pearl, 2000), in DMGs so-called *collider paths* (see Appendix B.3) can result in variables with an arbitrary distance to T in the graph being part of the Markov blanket of T. This can lead to an explosion in the size of the Markov blanket, which a user may want to control.

To that end, instead of enforcing (3.22), we allow for possible violations of the Markov blanket

condition (3.20) by introducing the error variables ξ_i, $i \in V \setminus \{T\}$, which yields the constraint:

$$m_i + \sum_{n \in N_{Ti}^I} v_i^n \geq 1 - \xi_i, \quad i \in V \setminus \{T\}. \tag{3.23}$$

Constraint (3.23) forces $\xi_i = 1$ for every $i \notin V(T)$ such that $i \not\perp\!\!\!\perp T | V(T)$; equivalently, $\xi_i = 1$ for each $i \notin V(T)$ that violates (3.20) in Definition 6. Our objective is to minimize violations of (3.20) while also controlling the total number of variables selected. We introduce a penalty constant $\lambda \in (0, 1)$ to modulate this trade-off, resulting in the following optimization problem:

$$\begin{aligned}
& \underset{\mathbf{m},\mathbf{v},\boldsymbol{\xi}}{\text{minimize}} && \sum_{i \in V \setminus \{T\}} ((1-\lambda) \cdot \xi_i + \lambda \cdot m_i) \\
\text{BlanketIP:} \qquad & \text{subject to} && (3.21) - (3.23), \\
& && v_i^n \in \{0,1\}, \quad n \in N_{Ti}^I,\ i \in V \setminus \{T\}, \\
& && m_i,\ \xi_i \in \{0,1\}, \quad i \in V \setminus \{T\}.
\end{aligned}$$

Let $(\mathbf{m}^*, \mathbf{v}^*, \boldsymbol{\xi}^*)$ be an optimal solution to BlanketIP, and let $V^*(T) = \{i \in V | m_i^* = 1\}$ be the corresponding set of variables. When λ is close to 1, BlanketIP emphasizes selecting a small number of variables; when λ is close to 0, the model emphasizes satisfaction of (3.20). Next, we present the main result of this section, which is that BlanketIP correctly retrieves the Markov blanket in the idealized setting where the independence tests are error-free.

Theorem 6. *Let Assumption 6 hold. If $\lambda \in (0, 1/|V|)$, then $V^*(T) = MB(T)$, i.e., the optimal solution to* BlanketIP *corresponds to the Markov blanket of T.*

Similar to Theorem 5, Theorem 6 can be interpreted as an asymptotic result, due to its dependence on Assumption 6. The intuition behind the small threshold of $1/|V|$ is that Theorem 6 also invokes Assumption 6, which corresponds to a setting where the independence tests have power

1 (due to an infinite sample size). In this high-powered setting, the trade-off between the size of $V^*(T)$ and satisfaction of (3.20) is non-existent, and so larger values of λ will needlessly penalize the size of the set of selected variables $V^*(T)$, resulting in inaccurate inference of the Markov blanket. However, in lower-powered settings, the trade-off between the size of $V^*(T)$ and satisfaction of (3.20) is non-trivial, making it possible for the most effective values of λ to be larger than $1/|V|$.

There are many variations of BlanketIP one could consider. For example, it may be fruitful to weight the nodes in the output blanket by the marginal or conditional dependence they have with T (as a measure of their predictivity), or introduce a constraint on the total number of nodes that the output may contain, especially if the causal discovery algorithm they are subsequently fed into is limited in its scalability. BlanketIP is intended to illustrate the utility of taking an integer optimization approach to Markov blanket-based feature selection.

3.5.3 Numerical Results

Our discussion of Markov blankets, including Definition 6 and the formulation of BlanketIP, has thus far been primarily in probabilistic terms. However, given the correspondence in Remark 1, we can naturally define the Markov blanket of a target node T within a causal graph as well (i.e., by replacing the independence condition in (3.20) with d-separation). This graphical interpretation of a Markov blanket allows us to numerically evaluate the performance of BlanketIP, namely, by checking how accurately it recovers the Markov blanket of a node in a ground-truth causal graph from observational data generated by the graph.

We conducted numerical experiments to examine the accuracy of BlanketIP, with a focus on the sensitivity of the output to the penalty λ. First, we generated 50 linear Gaussian models with $|V| = 10$ using three different sample sizes: $n \in \{250, 500, 1000\}$. Similar to §3.4.3, we set up this experiment using the code from Hyttinen et al. (2014). For each of the 50 instances, we

first found the *true* Markov blanket $MB(T)$ of a randomly selected target variable in the ground-truth graph (see Appendix B.3 for a characterization of Markov blankets for DMGs). Because Markov blanket discovery is akin to a binary classification problem (i.e., for each variable, determine whether it belongs to $MB(T)$), we measure the accuracy of the selected variables $V^*(T)$ returned by BlanketIP using *precision* and *recall* (Buckland and Gey, 1994). In our context, precision is the fraction of variables that are in both $V^*(T)$ and $MB(T)$ among the variables in $V^*(T)$, and recall is fraction of variables that are in both $V^*(T)$ and $MB(T)$ among the variables in $MB(T)$. Both performance metrics take on values between 0 and 1, with 1 representing perfect accuracy.

Table 3.3 reports the number of variables returned by BlanketIP, and the corresponding precision and recall for varying penalty constant λ and sample size n, averaged over 50 instances. As expected, Table 3.3 shows that the number of selected variables decreases with the penalty λ. Accordingly, precision and recall increase and decreases with λ, respectively. Further, Tables 3.3 shows that both performance metrics generally improve with the sample size n due to the increased power of the t-tests, although the effect of increasing the sample size from 250 to $1,000$ is subtle. These results demonstrate how the penalty term λ can be used to control the behavior of BlanketIP in a finite-sample setting with potential errors in the test results.

n/λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
250	5.52	5.33	4.51	3.41	2.58	1.78	1.44	1.00	1.00
500	6.02	5.64	4.82	3.56	2.60	2.22	1.76	1.18	1.00
1000	6.22	5.14	5.11	4.48	2.54	2.26	1.68	1.12	1.00
250	0.66	0.70	0.71	0.72	0.78	0.81	0.94	0.99	1.00
500	0.69	0.70	0.71	0.76	0.83	0.86	0.92	0.98	1.00
1000	0.73	0.74	0.77	0.78	0.86	0.88	0.94	0.99	1.00
250	0.93	0.92	0.84	0.67	0.60	0.50	0.38	0.32	0.30
500	0.89	0.89	0.83	0.71	0.62	0.53	0.44	0.35	0.33
1000	0.95	0.92	0.89	0.75	0.69	0.60	0.48	0.36	0.33

Table 3.3: Average number of selected variables $|V^*(T)|$ (top), precision (middle), and recall (bottom) for varying penalty λ and sample size n.

3.6 Conclusion

It is now well recognized that a sound understanding of causal relations is essential for effective decision-making and counterfactual prediction. In this chapter, we presented a new optimization-based method for *causal discovery*: the problem of learning causal relations from observational data. The key to our method is an iterative solution algorithm that makes use of conditional independence structure in the data to identify promising edges and paths to include in the output graph. Our method is one of the few in the literature that allows for both feedback cycles and unmeasured confounding, and performs favorably compared with those that do. Further, the computational tractability of our approach makes it a promising complement to existing empirical methods used for causal inference. Natural extensions to our approach include adding weights to the input conditions, relaxing the linear Gaussian assumption in the cyclic case, or evaluating our method's performance when we do not have access to the results of all possible conditional independence tests.

Chapter 4

Graphical Validation of Instrumental Variables

4.1 Introduction

Instrumental variables (Bowden and Turkington, 1990; Angrist et al., 1996; Angrist and Krueger, 2001) are one of the most widely used techniques for causal inference. Their primary use is in overcoming bias in estimates of causal effects that arise from unmeasured confounding. For example, consider the problem of estimating the causal effect of one variable, X, on another, Y. The correlation between X and Y may provide a misleading picture of the causal effect due to unmeasured confounders that influence both variables. Intuitively, a third variable, I, is called an *instrument* for X if it is correlated with X, and any effect I has on Y is exclusively via X. Under these conditions, any correlation between I and Y is then taken as evidence that X is a cause of Y.

A major challenge in using an instrumental variable is identifying an appropriate one in the first place. For the most part, instruments are selected subjectively based on domain knowledge. Angrist and Krueger (2001) write "good instruments often come from detailed knowledge of the economic

mechanism and institutions", and Imbens and Rosenbaum (2005) write "finding instruments is an art rather than a science". Although there is a recent line of work on developing tests for instrument validity under various assumptions (e.g., Kitagawa (2015); Mourifié and Wan (2017); Kédagni and Mourifié (2020)), the necessary conditions for valid instruments are often said to be unverifiable from data (Stock, 2002). This is because verifying whether a proposed instrument satisfies the critical *exclusion* criterion is generally difficult (Angrist et al., 1996; Stock, 2002). As a consequence, the use of an instrumental variable is typically accompanied by context-specific arguments in favor of its validity. The subjective nature of instrumental variables has also led to extensive work on the potential pitfalls of using weak instruments, and how to detect or overcome them (Bound et al. (1995); Staiger and Stock (1997); Stock et al. (2002); Murray (2006)).

In this chapter, we demonstrates how the methodology developed in Chapter 3 can be used to gain more insights on the validity of instrumental variables. We apply our method to US Census data used in the landmark study on the returns to education by Angrist and Krueger (1991) (hereafter, AK-91). We focus on AK-91 because it presents one of the pioneering applications of an *instrumental variable* and existing work on the instrument from AK-91 is extensive. Our goal is to show how our method can be used to investigate the validity of an instrumental variable within the framework of causal graphical models and causal discovery. Our findings are consistent with the literature on AK-91's instrument – that it is plausible, but potentially undermined by confounding.

4.2 Instrumental Variables and Graphical Criteria

Instrumental variable methods provide a powerful tool to address many issues that arise while inferring causation from observational data. Their main power is to enable causal inference even in settings where the treatment assignment is non-random. Even though the underlying assumptions for instrumental validity vary in the literature, three general assumptions can be identified for i to

be a valid *instrument* for the effect of j on k (Lousdal, 2018).

1. *The relevance assumption*: The instrument i has a causal effect on j.

2. *The exclusion restriction*: The instrument i affects the outcome k only through j.

3. *The exchangeability assumption*: The instrument i does not share common causes with the outcome k.

The following definition formalizes instrumental variables in a graphical setting:

Definition 7 (Instrumental variables)**.** *A node i is an instrument for the effect of j on k if* (i) $i \not\perp k|C$ *for* $C \subseteq V \setminus k$ *and* (ii) *every unblocked path from i to k contains an arrow pointing into j (Pearl, 2000).*

Note that while Definition 7(i) directly corresponds to the relevance assumption, the exclusion and exchangeability assumptions are captured through Definition 7(ii). Figure 4.1 provides examples of valid and invalid instruments given various *true* causal structures. In Figures 4.1(a) and (b), i is a valid instrument for estimating the effect of j on k because the causal relationship between i and k is always mediated by j. In Figure 4.1(c), i is an invalid instrument because the path $i \leftrightarrow l \rightarrow k$ violates Definition 7(ii).

Next, we show how our method can be used to intuitively examine the validity of an instrumental variable. Specifically, within our graphical framework, checking instrument validity amounts to learning causal structures over the relevant data and checking the extent to which the output graph abides by Definition 7.

4.3 Quarter-of-Birth Instrument from Angrist and Krueger (1991)

The causal effect of education on income is a classical question in economics with significant policy implications, but one that is challenging to measure due to unobserved confounders (Card, 1999).

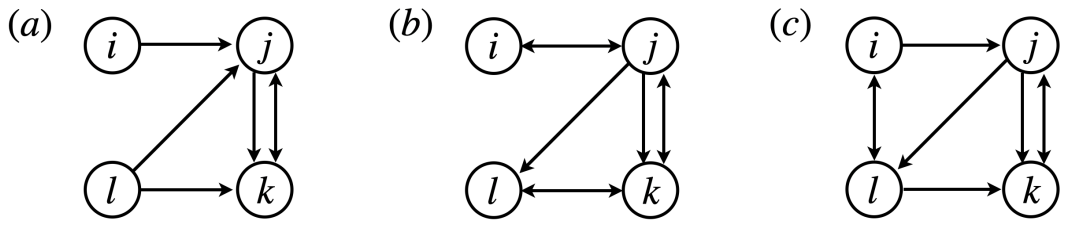


Figure 4.1: Node i is a valid instrument for the effect of j on k in (a) and (b), but an invalid instrument in (c) because the path $i \leftrightarrow l \rightarrow k$ violates Definition 7(ii).

The remedy for confounding proposed by AK-91 is to use *quarter-of-birth* as an instrument for years of education completed. The argument for this instrument is as follows: Because students are born year-round, the age at which students start school varies. Further, compulsory schooling laws in many states prohibit students from dropping out before they reach a certain age (e.g., their 16^{th} birthday). The combination of variability in starting ages and compulsory schooling laws effectively forces some students to complete more schooling than others, making quarter-of-birth correlated with education. Further, AK-91 argues that there is little reason to think quarter-of-birth would be correlated with income beyond its effect on education, and conclude it to be a valid instrument for estimating the effect of education on income.

AK-91's pioneering use of quarter-of-birth as an instrument for education has led to its adoption in numerous other studies (Buckles and Hungerman (2013)). Meanwhile, the validity of this instrument has been the subject of extensive debate and discussion (e.g., Bound et al. (1995), Card (1999), Staiger and Stock (1997), Angrist and Krueger (2001), Imbens and Rosenbaum (2005), Buckles and Hungerman (2013)). For quarter-of-birth to be a valid instrument for education, Definition 7 implies that it must be related to education (condition (i)), and that there cannot exist a path from quarter-of-birth to income that does not pass through education (condition (ii)). To that end, our approach will be to apply our method to the data from AK-91, and to check whether the quarter-of-birth variable indeed satisfies these two criteria.

4.4 Data and Experimental Setup

We focus on a subset of the data used in AK-91 containing information about 329,509 individuals taken from the 1980 US Census. Angrist and Krueger (1991) repeat their analysis for three cohorts separately – those born in the 1920s, 1930s, and 1940s – obtaining similar results across all three cohorts. For conciseness in our presentation, we use data from just the middle cohort. There are six available variables: QOB (quarter-of-birth, an integer value between 1 and 4), EDU (years of education completed), WAGE (weekly wage), RACE (race, 1 = Black) MAR (marital status, 1 = married), and SMSA (location of residence, 1 = Metropolitan Statistical Area). The data is publicly available (Angrist, 1991). To remove the effect of year-of-birth, we de-trended the data following the steps described in AK-91.

As discussed in Chapter 3, an important modeling decision in all constraint-based causal discovery methods, including ours, is the choice of the conditional independence test used to generate the input d-separation constraints. We use correlation-based t-tests for their simplicity and computational efficiency. Because the outcomes of the conditional independence tests depend critically on the significance level α used in the t-tests, we repeat our analysis for four different values: $\alpha \in \{0.05, 0.01, 0.001, 0.0001\}$.

Similar to other constraint-based methods, our algorithm does not assign confidence to the causal relations it uncovers – that is, an edge is either present or absent in the output. To introduce a notion of confidence in the discovered edges, we used a simple bootstrapping procedure. First, we re-sampled the data with replacement to construct 50 datasets, each having the same sample size as the original dataset. For each of the 50 repetitions, we apply the t-tests to generate the independence relations, which are input to EdgeGen to construct a causal graph. We terminate the algorithm if no improvement in the number of unsatisfied input constraints is found after three consecutive iterations. As a consequence of the bootstrap, our results take the form of relative edge

frequencies over the 50 graphs, where higher frequencies denote greater confidence in the causal relation implied by the edge.

4.5 Results

Figure 4.2 shows the four graphs obtained under each value of the significance level α, where the frequency of an edge is denoted by its thickness. The eight tables below correspond to the four graphs in Figure 4.2, separated into directed and bi-directed edges. All edge frequencies are normalized, where 1 indicates that the edge appeared in the output graph of all 50 bootstrap repetitions.

Note that edge frequencies generally increase in the significance level α, which is expected. Intuitively, this occurs because a rejection of the null hypothesis in each of the conditional independence tests implies that the variables in question are conditionally dependent (and thus conditionally d-connected), and satisfying each d-connection relation requires a path in the output graph. As a result, smaller values of α generate fewer d-connection relations for the input of our algorithm, yielding sparser graphs.

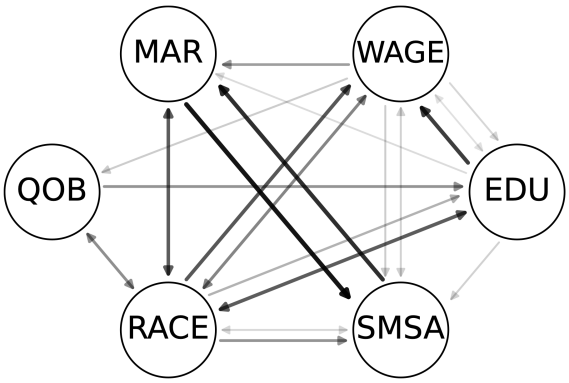
We use our results to address three questions that determine the validity of quarter-of-birth as an instrument for education:

(Q1) Is there a relationship between QOB and EDU?

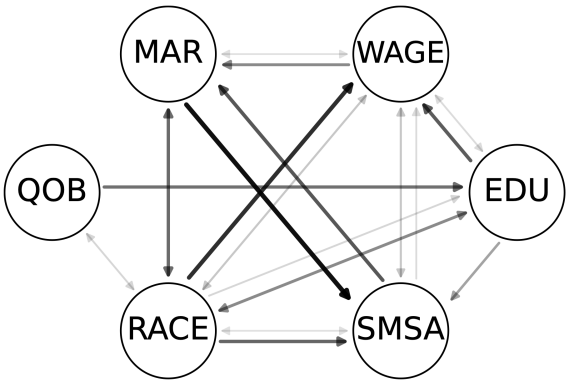
(Q2) Is there a (potentially latent) confounder between QOB and WAGE?

(Q3) Is QOB a direct cause of WAGE?

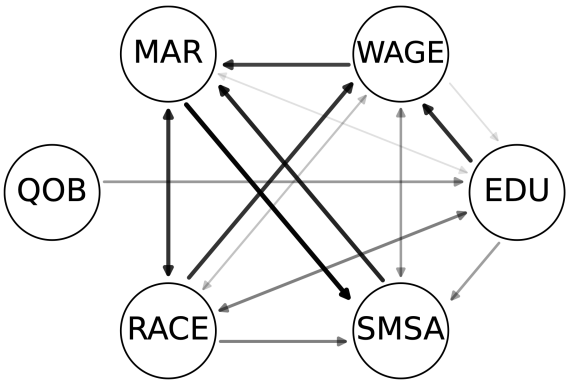
If the answer to the first question is yes, then condition (i) in Definition 7 is satisfied. If the answer to either the second or third question is yes, then condition (ii) is violated. Note that QOB need



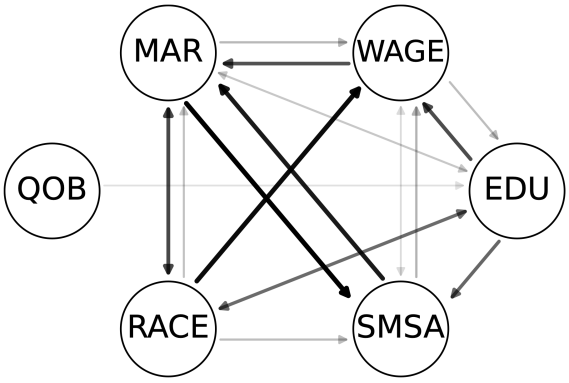
((a)) $\alpha = 0.05$.



((b)) $\alpha = 0.01$.



((c)) $\alpha = 0.001$.



((d)) $\alpha = 0.0001$.

Figure 4.2: Edge frequencies over 50 bootstrap repetitions of EdgeGen applied to US Census data from Angrist and Krueger (1991).

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.76	0.14	0	0	0.18
WAGE	0.16	-	0.38	0.20	0	0.20
MAR	0.08	0.06	-	0	0	0.92
QOB	0.42	0	0	-	0	0
RACE	0.30	0.66	0.04	0	-	0.42
SMSA	0.10	0.08	0.82	0	0	-

Table 4.1: Directed edge frequency for $\alpha = 0.05$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.64	0.08	0	0	0.36
WAGE	0.10	-	0.46	0.04	0	0.08
MAR	0.02	0.06	-	0	0	0.94
QOB	0.56	0.04	0.02	-	0	0
RACE	0.16	0.82	0.08	0	-	0.60
SMSA	0.04	0.14	0.66	0	0	-

Table 4.2: Directed edge frequency for $\alpha = 0.01$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.82	0.06	0	0	0.38
WAGE	0.12	-	0.74	0.02	0	0.06
MAR	0.04	0.10	-	0	0	0.98
QOB	0.40	0	0	0	-	0
RACE	0.08	0.80	0.08	0	-	0.50
SMSA	0.04	0.10	0.86	0	0	-

Table 4.3: Directed edge frequency for $\alpha = 0.001$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.70	0	0	0	0.60
WAGE	0.26	-	0.68	0	0	0
MAR	0.02	0.28	-	0	0	1.00
QOB	0.12	0	0	-	0	0
RACE	0.04	0.96	0.24	0	-	0.26
SMSA	0	0.30	0.90	0	0	-

Table 4.4: Directed edge frequency for $\alpha = 0.0001$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.12	0.04	0.08	0.64	0.06
WAGE	0.12	-	0.10	0.06	0.48	0.18
MAR	0.04	0.10	-	0	0.64	0.06
QOB	0.08	0.06	0	-	0.44	0
RACE	0.64	0.48	0.64	0.44	-	0.16
SMSA	0.06	0.18	0.06	0	0.16	-

Table 4.5: Bi-directed edge frequency for $\alpha = 0.05$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.14	0.06	0.02	0.46	0.02
WAGE	0.14	-	0.12	0.04	0.22	0.22
MAR	0.06	0.12	-	0.02	0.58	0.02
QOB	0.02	0.04	0.02	-	0.16	0.02
RACE	0.46	0.22	0.58	0.16	-	0.12
SMSA	0.02	0.22	0.02	0.02	0.12	-

Table 4.6: Bi-directed edge frequency for $\alpha = 0.01$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.02	0.12	0.02	0.50	0.06
WAGE	0.02	-	0.04	0	0.24	0.38
MAR	0.12	0.04	-	0	0.80	0.04
QOB	0.02	0	0	-	0.06	0
RACE	0.50	0.24	0.80	0.06	-	0.10
SMSA	0.06	0.38	0.04	0	0.10	-

Table 4.7: Bi-directed edge frequency for $\alpha = 0.001$.

	EDU	WAGE	MAR	QOB	RACE	SMSA
EDU	-	0.04	0.22	0.02	0.58	0.10
WAGE	0.04	-	0	0	0.06	0.14
MAR	0.22	0	-	0	0.72	0.08
QOB	0.02	0	0	-	0.02	0
RACE	0.58	0.06	0.72	0.02	-	0
SMSA	0.10	0.14	0.08	0	0	-

Table 4.8: Bi-directed edge frequency for $\alpha = 0.0001$.

not be a cause of EDU for condition (i) of Definition 7 to hold, and that any association between the variables is sufficient (Angrist et al., 1996; Pearl, 2000).

Q1 is straightforward to answer even without investigating causal structures – one can simply check the linear correlation between QOB and EDU, which reveals a weak association. Still, it is instructive to consider how the relationship between QOB and EDU is captured by our graphical framework. Our results show that the edge QOB $\rightarrow$ EDU appears with a frequency of 0.40–0.56 for $\alpha \in \{0.05, 0.01, 0.001\}$, indicating that condition (i) in Definition 7 is not satisfied in half of the output graphs. For $\alpha = 0.0001$, the frequency of QOB $\rightarrow$ EDU drops to 0.12. The fragility of this edge, even at larger values of α and despite the large sample size, speaks to the weakness of QOB as an instrument for EDU.

The weak association between quarter-of-birth and education is, on its own, not problematic with respect to quarter-of-birth's validity as an instrument. The concern here is that because the relationship between QOB and EDU is weak, then even a weak relationship between QOB and WAGE that bypasses EDU can lead to a large bias in the estimates of the causal effect of education on income (Bound et al., 1995). This brings us to Q2. Notably, our results provide evidence of confounding between QOB and RACE: The bi-directed edge QOB $\leftrightarrow$ RACE appears with a frequency of 0.44 for $\alpha = 0.05$, although it vanishes at smaller values of α. While QOB $\leftrightarrow$ RACE only appears at larger values of α, the presence of this edge threatens the validity of quarter-of-birth as an instrument for education. Because the edge RACE $\rightarrow$ WAGE is quite robust (appearing with frequency 0.66-0.96 depending on α), the edge QOB $\leftrightarrow$ RACE violates condition (ii) in Definition 7 by creating a causal path from QOB to WAGE that does *not* pass through EDU.

With respect to Q3, our results show that the frequency of an edge (either directed or bi-directed) between QOB and WAGE is at most 0.06 (under $\alpha = 0.05$), and precisely 0 for $\alpha \in \{0.001, 0.0001\}$. These low frequencies suggest that the instrument does not materially suffer from

QOB being a direct cause of WAGE, or from latent confounding between QOB and WAGE.

Lastly, although our focus has been on quarter-of-birth, our results suggest that education is indeed a cause of income: The edge EDU $\rightarrow$ WAGE persists across all four values of α, and appears with frequency 0.7 for $\alpha = 0.0001$. At this significance level, only four other edges appear with greater frequency: MAR$\rightarrow$ SMSA (1.00), RACE $\rightarrow$ WAGE (0.96), SMSA $\rightarrow$ MAR (0.90), and RACE $\leftrightarrow$ MAR (0.72).

4.5.1 Comparison with Bound et al. (1995) and Buckles and Hungerman (2013)

Two of the most well-known critiques of the quarter-of-birth instrument from AK-91 are presented by Bound et al. (1995) (hereafter BJB-95) and Buckles and Hungerman (2013) (BH-13). As a point of reference, we discuss how our results compare to statements made in BJB-95 and BH-13.

The main criticism in BJB-95, which is related to our Q1 above, is that quarter-of-birth's association with education is so weak that even minimal confounding may lead to biased estimates, despite the large sample sizes in AK-91. As noted above, this weak association between quarter-of-birth and education is reflected in our results by the edge QOB $\rightarrow$ EDU vanishing at smaller values of α, while other edges persist.

More interestingly, BJB-95 also suggest that quarter-of-birth may be associated with family characteristics that are predictive of an individual's income (Q2). In particular, they argue that race may be associated with quarter-of-birth, and also point to research that finds families with high incomes are less likely to have children in the winter months (Kestenbaum, 1987). These conjectures about potential confounding due to family background are rigorously examined by BH-13, who propose that maternal characteristics can explain a significant share of the association between quarter-of-birth and income. Using birth certificate and US Census data, BH-13 find that children born in the winter are more likely to have mothers that are non-white, teenagers, and

lacking a high school diploma. BH-13 state that because quarter-of-birth is associated with family background (which is itself related to income), the quarter-of-birth instrument violates the critical *exclusion* criterion. Notably, our detection of a causal relationship between RACE and QOB is well-aligned with both BJB-95 and BH-13's claims of potential confounding between quarter-of-birth and income due to race.

Beyond the effect of race, BJB-95 also argue that it is plausible that quarter-of-birth has a direct effect on income, and point to research in psychology and education for possible mechanisms (Q3). Our results do not support this claim, because we do not find quarter-of-birth to be a direct cause of income, nor do we detect any confounding between those variables other than race. This is perhaps not surprising, given that the BJB-95 simply argue for the plausibility of quarter-of-birth having a direct effect on income, but also state that they are not aware of any "indisputable evidence" that such an effect exists. Our results suggest that any confounding between QOB and WAGE is for the most part already captured by the variable RACE.

We highlight a few limitations of our study. As discussed above, the output graphs are sensitive to the choice of conditional independence test and the significance level used. We tested sensitivity to α, but not to the choice of independence test; the use of other tests may alter our results. Further, there may exist multiple graphs within the Markov equivalence class of graphs implied by the input independence relations, although our method seeks to return only one of them. Lastly, we terminate the algorithm after three consecutive iterations with no improvement in the number of satisfied input constraints, and adjusting this criterion may also affect our results.

4.6 Conclusion

In this chapter, we demonstrate how the method proposed in Chapter 3 can be used to investigate validity of instrumental variables. We apply our method to US Census data used in the landmark

study on the returns to education by Angrist and Krueger (1991) (hereafter, AK-91) as it presents one of the pioneering applications of an *instrumental variable*. The agreement between the literature and the causal structures uncovered by our algorithm suggests that a causal discovery approach can be a useful complement to existing empirical methods. Further, the computationally efficiency of our proposed method makes it more implementable in practice compared to extant causal discovery techniques. Although our focus in this section was to examine the validity of an instrumental variable, our method can also be used more generally to uncover causal structure in data – for example, identifying new causal relations in observational data can help point us toward potentially fruitful directions for further research.

Our analysis is not necessarily unique to our proposed method. In principle, one could repeat the validation exercise presented here using other causal discovery techniques, including those that restrict the search space to DAGs, which are extremely efficient. However, because our approach allows us to detect latent confounding, it can provide additional insight into the causal relations among variables of interest that DAG-based approaches cannot.

Chapter 5

Conclusions

In this thesis, we have developed methodologies to uncover the underlying mechanisms that are governing the effects of strategic decisions using optimization, economic modeling, and statistics. We briefly summarize the key contributions of each chapter here.

In Chapter 2, we developed an integer programming formulation that is based on principal-agent models to analyse the effects of incentive contracts in online labor platforms. We proposed an intuitive estimator for a general class of principal-agent models where agent actions are hidden. Our agent model is non-parametric, and imposes minimal structural assumptions on the agent's utility function. We showed that the estimator is identifiable and statistically consistent under some mild assumptions. Since solving the exact estimator using off-the-shelf optimization solvers is non-viable for larger instances, we propose an approximation scheme based on discretizing the original parameter space. We then developed an iterative solution algorithm in the spirit of column generation that exploits statistical properties of the integer programming formulation. We showcase the efficacy of the estimator in two sets of numerical experiments. First, we use synthetic instances to show that the proposed algorithm produces solutions that are competitive with the exact estimator in a fraction of the computational time. To demonstrate our method's applicability

in real world, we conducted an experiment on a crowd-work platform.

In Chapter 3, we considered how to learn causal relations from observational data using optimization. Our main contribution is a new optimization-based method for causal discovery. We consider an extremely general search space where feedback cycles and unobserved confounders are allowed. To achieve better scalability over this large search space, we proposed an iterative solution approach that exploits the conditional (in)dependence structure in the data to detect "promising" candidate edges in the underlying graph, which are then assembled into a causal graph by an optimization model. We computationally showed that our approach performs better in terms of scalability compared to state-of-the-art methods.

In Chapter 4, we focused on investigating the validity of an instrumental variable using the method developed in Chapter 3. We applied our method to the well-known dataset on educational attainment and income from Angrist and Krueger (1991), which contains one of the most influential applications of an instrumental variable, but one whose validity has been the subject of debate. The agreement between the existing literature and the causal structures uncovered by our algorithm suggests that causal discovery can be a useful complement to existing empirical methods.

Appendices

Appendix A

Estimating Effects of Incentive Contracts in Online Labor Platforms

A.1 Identifying an Optimal Incentive Contract

A natural question is whether it is possible to use our estimated model to identify an optimal incentive contract from the set R. Here, we establish an important property of our model: our agent-based method for estimating the mapping from contracts to outcomes yields an optimal contracting formulation that is simple and tractable. This tractability is a direct consequence of our specification of the agent model, and is not guaranteed if alternative methods are used to estimate the mapping from contracts to outcomes.

Let $\zeta_j(\mathbf{r})$ be the utility (e.g., of a principal) under outcome $\{\xi = j\}$ and contract $\mathbf{r}$, and suppose we are interested in identifying a contract $\mathbf{r} \in R$ that maximizes expected utility: $U(\mathbf{r}) = \sum_{j \in J} \Pr(\xi = j|\mathbf{r})\zeta_j(\mathbf{r})$. As we have assumed throughout, the distribution $\Pr(\xi = j|\mathbf{r})$ may be unknown in practice, which implies the utility function $U(\mathbf{r})$ is also unknown. However, given data on past contracts and outcomes, a reasonable approximation is to first estimate $\hat{\Pr}(\xi = j|\mathbf{r})$ for all

$\mathbf{r} \in R$, which produces an estimate of U:

$$\hat{U}(\mathbf{r}) = \sum_{j \in J} \hat{\Pr}(\xi = j|\mathbf{r})\zeta_j(\mathbf{r}).$$

Then, a sensible approximation is to find a contract in R that maximizes the estimated expected utility:

$$(\mathsf{OC}) \qquad \underset{\mathbf{r} \in R}{\text{maximize}} \ \hat{U}(\mathbf{r}).$$

Next, suppose $\zeta_j(\mathbf{r})$ is convex in $\mathbf{r}$, and let R be a convex set. Then, the tractability of the optimal contracting problem (OC) depends critically on the expression for $\hat{\Pr}(\xi = j|\mathbf{r})$, which in turn is determined by the method used to estimate it.

A useful structural property of our approach is that the expression for $\hat{\Pr}(\xi = j|\mathbf{r})$ makes problem OC quite straightforward. Specifically, within our modeling framework, we have $\hat{\Pr}(\xi = j|\mathbf{r}) = \hat{\pi}_{\hat{a}(\mathbf{r}),j}$, where $\hat{a}(\mathbf{r}) \in \text{argmax}_{a \in A} \left\{ \sum_{j \in J} \hat{\pi}_{aj} r_j - c_a \right\}$. Then, for a given estimate $\hat{\boldsymbol{\pi}}$, we can write the optimal contracting problem as:

$$\begin{aligned}
& \underset{\mathbf{r}}{\text{maximize}} && \hat{\pi}_{\hat{a}(\mathbf{r}),j}\zeta_j(\mathbf{r}) && \text{(A.1a)} \\
(\mathsf{OC}(\hat{\boldsymbol{\pi}})) \qquad & \text{subject to} && \hat{a}(\mathbf{r}) \in \text{argmax}_{a \in A} \sum_{j \in J} \hat{\pi}_{aj} r_j - c_a, && \text{(A.1b)} \\
& && \mathbf{r} \in R. && \text{(A.1c)}
\end{aligned}$$

This formulation leads to the following result, which we present without proof:

Lemma 2. *If R is a convex set and $\zeta_j(\mathbf{r})$ is convex in $\mathbf{r}$, then an optimal solution to $\mathsf{OC}(\hat{\boldsymbol{\pi}})$ can be obtained by solving $|A|$ convex problems.*

To see why the lemma holds, consider the following formulation for a fixed action a:

$$\underset{\mathbf{r}}{\text{maximize}} \quad \sum_{j \in J} \hat{\pi}_{aj} \zeta_j(\mathbf{r}) \tag{A.2a}$$

$$\text{subject to} \quad \sum_{j \in J} r_j \hat{\pi}_{aj} - c_a \geq \sum_{j \in J} r_j \hat{\pi}_{bj} - c_b, \quad b \in A, \tag{A.2b}$$

$$\mathbf{r} \in R. \tag{A.2c}$$

The subproblem (A.2) finds a utility-maximizing contract from $R_a(\hat{\boldsymbol{\pi}})$ – the subset of R such that action a is optimal for the agent under $\hat{\boldsymbol{\pi}}$. Specifically, the objective (A.2a) gives the expected payoff under contract $\mathbf{r}$ and action a, the constraint (A.2b) restricts the contracts $\mathbf{r}$ to those that make action a optimal for the agent, and constraint (A.2c) restricts $\mathbf{r}$ to the set R. The subproblem (A.2) is clearly convex if R is a convex set and $\zeta_j(\mathbf{r})$ is convex in $\mathbf{r}$. Therefore, $\mathsf{OC}(\hat{\boldsymbol{\pi}})$ can be easily solved by solving the subproblem (A.2) once for each action $a \in A$, and selecting the action and corresponding contract that maximizes (A.2a). Note that (A.2) may be infeasible if $R_a(\hat{\boldsymbol{\pi}})$ is empty (i.e., there exists an action such that no contract makes it optimal), but that at least one $R_a(\hat{\boldsymbol{\pi}})$ must be non-empty, so it is always possible to solve $\mathsf{OC}(\hat{\boldsymbol{\pi}})$. In summary, once equipped with an estimate $\hat{\boldsymbol{\pi}}$, our model allows us to easily optimize over the set R to identify an optimal contract.

As a point of comparison, note that the general form optimal contracting problem OC may not be straightforward to solve if an alternative method is used to estimate $\hat{\Pr}(\xi = j|\mathbf{r})$. For example, in the context of multinomial logistic regression (where the contract vector $\mathbf{r}$ are the independent variables and the outcomes $j \in J$ are classes), $\hat{\Pr}(\xi = j|\mathbf{r})$ is given by the logit function, which is non-convex in $\mathbf{r}$, making OC non-trivial to solve. Similarly, in non-parametric classification methods such as classification trees, there may not exist a closed form expression for $\hat{\Pr}(\xi = j|\mathbf{r})$, in which case solving OC is far from straightforward.

A.2 Additional Computational Results

A.2.1 Prediction Error Comparison with Multinomial Logistic Regression and Classification Trees

In this section, we further examine the predictive performance of the estimator using synthetic data. In particular, we assess the predictive performance of PA-D+ and the benchmark methods under different specifications of the underlying data generation process. The setup is as follows: For each observation $i \in I$, we construct $\mathbf{r}^i$ by drawing d values from the standard uniform distribution and sorting them in ascending order, so that $r^i_j \leq r^i_{j+1}$ for $j = 1, \ldots, d-1$, which reflects the notion that higher outcomes should correspond to higher payments. For PA-D+, action costs are given by $c_a = a/m$ for $a = 1, \ldots, m$. We then simulate the outcome data from discrete distributions of the form

$$\Pr(\xi = j|\mathbf{r}) = \frac{g_j(\mathbf{r})}{\sum_{k=1}^{d} g_k(\mathbf{r})},$$

where $g_j(\mathbf{r})$, $j \in J$ are functions that determine the outcome distribution under a contract $\mathbf{r}$. We test different underlying models by considering $g_j(\mathbf{r}) = r_j$, $g_j(\mathbf{r}) = \sqrt{r_j}$ and $g_j(\mathbf{r}) = (1 + r_j^2)$ (a natural interpretation is that we are testing predictive performance under different "ground-truth" models). For each of these three models, we consider six problem sizes: $(m, d, n) \in \{(2, 2, 100), (2, 2, 1000), (4, 5, 100), (4, 5, 1000), (5, 10, 100), (5, 10, 1000)\}$. For each of these six instances, we again fit all three models to a bootstrap sample of size n, and measure predictive performance on the out-of-bootstrap sample, for 100 repetitions. The prediction error is measured using the mean absolute error (MAE) given in (2.16). In a second set of experiments, we add a noise term ϵ to $g_j(\mathbf{r})$, where ϵ is drawn independently from the standard uniform distribution for each observation $i \in I$.

Tables A.1 and A.2 show the average prediction errors over 100 bootstrap repetitions with and

without the additional noise term, respectively. Minimum prediction errors are shown in bold. Table A.1 shows that all three methods have comparable performance when there is no noise term. In contrast, Table A.2 shows that when the noise term is included to generate the outcome data, the predictive performance of MLR degrades considerably, whereas PA-D+ performs well across all instances. Intuitively, this difference in performance is due to the non-parametric nature of PA-D+, which makes predictions based on empirical distributions found within each bootstrap sample, allowing it to fit the data well. We also find that classification trees are competitive in several instances in Table A.2, which is unsurprising given their flexibility.

Interestingly, PA-D+ shows little evidence of overfitting. This may be because the estimator searches over the restricted parameter set $\widetilde{\Pi}$, which is constructed from the data, instead of the full space of possible distributions Π. While this restriction on the parameter space is motivated by computational tractability, it may also help the model avoid overfitting by forcing the estimator to only choose among a finite set of empirical distributions that appear in the data (see Algorithm 2). Lastly, it is worth noting that all three methods may be tuned further; for example, one can adjust the maximum number of splits in the classification trees, or adjust the parameters ρ, α and S in PA-D+; doing so may produce different results. Nonetheless, given that we do not exhaustively tune PA-D+ to the data, Tables A.1 and A.2 suggest that PA-D+ performs favorably compared to well-known prediction methods, and is robust to the underlying data generation process.

A.2.2 Solution Time Comparison with Maximum Likelihood Estimation

Naturally, one might ask whether $\boldsymbol{\pi}^0$ can also be estimated through a maximum likelihood estimation (MLE) approach. Here, we present numerical examples to show that an MLE approach may not be tractable for the non-parametric agent model that we study in this paper, due to the potentially large search space represented by the parameter set Π and the nonconvexity of the

$g_j(\mathbf{r})$	m	d	n	PA-D+	MLR	CT
r_j	2	2	100	**0.08**	0.09	0.09
	2	2	1000	0.03	0.03	**0.02**
	4	5	100	**0.07**	0.08	**0.07**
	4	5	1000	**0.02**	0.03	**0.02**
	5	10	100	**0.04**	0.06	0.05
	5	10	1000	**0.02**	0.02	**0.02**
$\sqrt{r_j}$	2	2	100	0.09	**0.03**	0.11
	2	2	1000	**0.03**	**0.03**	**0.03**
	4	5	100	**0.06**	**0.06**	**0.06**
	4	5	1000	**0.02**	**0.02**	**0.02**
	5	10	100	**0.04**	0.05	0.05
	5	10	1000	0.02	0.02	**0.01**
$(1 + r_j)^2$	2	2	100	**0.08**	0.09	0.09
	2	2	1000	0.03	**0.02**	0.03
	4	5	100	**0.07**	0.08	**0.07**
	4	5	1000	**0.02**	0.03	**0.02**
	5	10	100	**0.05**	**0.05**	0.06
	5	10	1000	**0.02**	**0.02**	0.02

Table A.1: Out-of-bootstrap prediction errors of PA-D+, multinomial logistic regression (MLR) and classification trees (CT) under varying data generation processes (without noise), averaged over 100 bootstrap repetitions. Minimum errors are bolded.

$g_j(\mathbf{r})$	m	d	n	PA-D+	MLR	CT
$r_j + \epsilon$	2	2	100	**0.07**	0.20	0.25
	2	2	1000	**0.02**	0.34	0.22
	4	5	100	**0.06**	0.28	0.10
	4	5	1000	**0.02**	0.22	**0.02**
	5	10	100	**0.04**	0.16	0.05
	5	10	1000	**0.02**	0.13	**0.02**
$\sqrt{r_j} + \epsilon$	2	2	100	**0.11**	0.33	0.32
	2	2	1000	**0.03**	0.35	0.20
	4	5	100	**0.07**	0.27	0.08
	4	5	1000	**0.02**	0.22	**0.02**
	5	10	100	**0.05**	0.15	0.06
	5	10	1000	**0.02**	0.13	**0.02**
$(1 + r_j)^2 + \epsilon$	2	2	100	**0.10**	0.19	0.31
	2	2	1000	**0.03**	0.23	0.20
	4	5	100	**0.06**	0.30	0.11
	4	5	1000	**0.02**	0.25	0.03
	5	10	100	**0.05**	0.15	**0.05**
	5	10	1000	**0.02**	0.13	**0.02**

Table A.2: Out-of-bootstrap prediction errors of PA-D+, multinomial logistic regression (MLR) and classification trees (CT) under varying data generation processes (with noise), averaged over 100 bootstrap repetitions. Minimum errors are bolded.

likelihood function.

Note that for each $i \in I$, ξ^i is the outcome in J that was observed under $\mathbf{r}^i$. Then, based on the agent model (2.1), the log-likelihood of seeing the data $(\mathbf{r}^i, \xi^i)$, $i \in I$ under the parameter $\boldsymbol{\pi}$ is

$$\tilde{L}(\boldsymbol{\pi}) = \sum_{i \in I} \log(\pi_{a,\xi^i}), \text{ where } a = \underset{a \in A}{\operatorname{argmax}} \left\{ \sum_{j \in J} \pi_{aj} r^i_j - c_a \right\}. \tag{A.3}$$

One challenge with maximizing the log-likelihood function $\tilde{L}(\boldsymbol{\pi})$ is that it is discontinuous at values of $\boldsymbol{\pi}$ where two or more actions are optimal. Intuitively, this is because for these values of $\boldsymbol{\pi}$, a small change in the parameter $\boldsymbol{\pi}$ can make the agent's optimal action "jump" from one action to another, which leads to a discontinuity in $\tilde{L}(\boldsymbol{\pi})$. Therefore, non-linear optimization techniques that require the likelihood function to be continuous and differentiable cannot be used to maximize $\tilde{L}(\boldsymbol{\pi})$.

An intuitive approach to optimizing $\tilde{L}(\boldsymbol{\pi})$ that does not require differentiability of $\tilde{L}(\boldsymbol{\pi})$ is to perform an exhaustive grid search over the parameter set Π. We set up the following numerical experiment to test the viability of this method. First, we randomly generate problem data using the process described in §2.4.4, where $(m, d) \in \{(2,2), (2,3), (2,4), (2,5), (3,2), (3,3), (3,4)\}$, and $n \in \{100, 500, 1000\}$. We then discretize the parameter set Π in increments of 0.1 for each element (a, j) in the matrix $\boldsymbol{\pi}$. For example, in the instances where $d = 2$, we search over the set $\prod_{a \in A} P_{\text{grid}}$, where $P_{\text{grid}} = \{(0,1), (0.1, 0.9), (0.2, 0.8), \ldots, (0.9, 0.1), (1,0)\}$. We then evaluate the log-likelihood function $\tilde{L}(\boldsymbol{\pi})$ for each value of $\boldsymbol{\pi}$ in the grid, and set the estimate as the solution that yields the largest value of $\tilde{L}(\boldsymbol{\pi})$. We set a time limit of 3600 CPU seconds for each instance.

Table A.3 summarizes the results of this search procedure, averaged over 10 repetitions per problem size. The results indicate that for even small instances, an exhaustive search may require up to one CPU hour, making this method impractical for larger problem sizes. This is an unsurprising

			MLE	
m	d	n	Time	Error
2	2	100	0	0.11
2	2	500	0	0.05
2	2	1000	1	0.04
2	3	100	8	0.07
2	3	500	14	0.05
2	3	1000	22	0.04
2	4	100	144	0.07
2	4	500	264	0.05
2	4	1000	424	0.05
2	5	100	1751	0.08
2	5	500	3301	0.05
2	5	1000	-	-
3	2	100	3	0.09
3	2	500	5	0.08
3	2	1000	7	0.08
3	3	100	488	0.08
3	3	500	931	0.07
3	3	1000	1471	0.07
3	4	100	-	-
3	4	500	-	-
3	4	1000	-	-

Table A.3: Solution time (CPU seconds) and normalized estimation error of MLE via grid search, averaged over 10 trials. Number of actions, outcomes and sample size are denoted by m, d, and n, respectively. Instances that did not solve to optimality under 3600 CPU seconds are omitted when calculating average estimation error. Dashes indicate no instance solved to optimality within 3600 CPU seconds in any trial.

result, given that the number of solutions to be evaluated grows exponentially in the number of actions, m, and the number of outcomes, d. It can be observed from the results in Table A.3 that our estimation procedure obtains comparable estimation errors in approximately 1 CPU minute for similar problem sizes. We note here that other heuristics for searching over Π (e.g. genetic algorithms) may be more fruitful than a simple grid search, although they will also be subject to a search space where the number of grid points grows exponentially in m and d.

A.3 Agent Heterogeneity

The estimation procedure developed in Chapter 2 can be naturally extended to accommodate heterogeneous agents. We model heterogeneity non-parametrically by assuming each agent has a categorical *type* that is observable from the data, and allow an agent's action costs and outcome distributions to depend on the agent's type.

In this section, we show that our main estimator PA can be modified to incorporate agent heterogeneity, and that consistent estimation remains possible in the heterogeneous setting. In particular, we present an analogous result to Theorem 1 for an estimator that accounts for agent heterogeneity. Note that similar results to Proposition 2 and Theorem 3 can also be obtained under agent heterogeneity, by applying similar arguments. For conciseness, we will only formally present an analogue to Theorem 1 under agent heterogeneity.

We also show how *a priori* information about the relative efficiency of different agent types can be incorporated into our estimator. Note that because agent types are observable, an intuitive procedure is to simply segment the data according to agent type, and then apply the base estimator PA-D+ to each segment separately. However, if additional information about the relative performance of different agent types is available, then one might expect an estimator that *pools* the data across all agent types to outperform a naive application of PA-D+ to each type separately. We present numerical results that show that incorporating this additional efficiency information can indeed improve estimation accuracy when data is limited, but can degrade accuracy when data is abundant, due to our approximation scheme.

The extension of our model to heterogeneous agents is intuitive; therefore, the development in this section will closely follow §2.2 and §2.3. In §A.3.1, we present the estimator formulation in the presence of agent heterogeneity, and discuss how *a priori* information about the relative efficiency of different agent types can be embedded in the parameter set Π. In §A.3.2, we present a corollary

to Theorem 1 that shows consistent estimation is possible under agent heterogeneity. In §A.3.3, we present the optimization model and column generation algorithm used to produce an analogue to the base estimator PA-D+, and provide numerical examples.

A.3.1 Estimator

Suppose that in addition to observing the contract $\mathbf{r}^i$ and outcome ξ^i, we also observe the agent's *type*, θ^i. Agent types are categorical and indexed by the set K. Thus, each historical observation consists of the triple $(\mathbf{r}^i, \xi^i, \theta^i)$. With a slight abuse of notation, we define the agent model using the parameter $\boldsymbol{\pi} \in \mathbb{R}_+^{m\times d\times|K|}$, where π_{aj}^k is the probability that action a leads to outcome j when taken by a type k agent. As before, we let $\boldsymbol{\pi}^0$ denote the true model parameter to be estimated, where $\Pi \subseteq \mathbb{R}_+^{m\times d\times|K|}$ is the parameter set and $\boldsymbol{\pi}^0 \in \Pi$. Next, let c_a^k be the cost of taking action a for a type k agent, and let $\mathbf{c}^k$ denote a type k agent's cost vector. We assume all agents have the same outcome set, A.[1] Let $A_k(\mathbf{r}^i, \boldsymbol{\pi}) = \underset{a\in A}{\operatorname{argmax}}\{\sum_{j\in J} \pi_{aj}^k r_j^i - c_a^k\}$ denote the set of optimal actions for a type k agent under the contract $\mathbf{r}^i$, and define $I^k = \{i \in I | \theta^i = k\}$ as the subset of observations where the agent is type k. The loss function $L_n^k(\boldsymbol{\pi})$ for a type k agent is then

$$L_n^k(\boldsymbol{\pi}) = \underset{x,\boldsymbol{\eta},\boldsymbol{\omega}}{\text{minimize}} \;\; \|\boldsymbol{\pi} - \boldsymbol{\omega}\|_1 \tag{A.4a}$$

$$\text{subject to} \;\; x^i \in A_k(\mathbf{r}^i, \boldsymbol{\pi}), \quad i \in I^k, \tag{A.4b}$$

$$\omega_{aj} = \frac{1}{|\{i \in I^k | x^i = a\}|} \sum_{i\in\{i\in I^k | x^i = a\}} y_j^i, \quad a \in A,\; j \in J. \tag{A.4c}$$

Note that the loss function (A.4) is similar to the loss function provided for a single agent type in §2.2.2, but is defined over the subset of observations I^k instead of I. The estimate of $\boldsymbol{\pi}^0$ is then attained at a minimizer of the sum of loss functions over Π:

[1] The assumption that all agent types share the same set of actions is without loss of generality; an action a for agent k, we may assume $c_a^k = \infty$.

$$\text{(PA-H)} \qquad \hat{\boldsymbol{\pi}}_n = \operatorname*{argmin}_{\pi \in \Pi} \sum_{k \in K} L_n^k(\boldsymbol{\pi}).$$

Note that PA-H is not necessarily equivalent to solving PA for each agent type separately, because the parameter set Π may link estimation problem across agent types based on *a priori* information. In particular, if we have additional information about the relative performance of different agent types, then pooling the data and solving PA-H effectively allows us to use data generated by one agent type to estimate the parameters for another. Next, we provide two examples of how *a priori* information can be incorporated into the model.

Outcome efficiency.

It is common in incentive problems for agents to be heterogeneous with respect to some notion of *efficiency* (Laffont and Martimort, 2009). Here we illustrate how different definitions of agent efficiency can be captured within our framework. Without loss of generality, assume outcomes are ordered based on the preferences of the principal, so that outcome $j+1$ is preferred to outcome j, for $j = 1, \ldots, d-1$. Suppose also that all agents share the same action costs ($\mathbf{c}^1 = \mathbf{c}^2 = \ldots, \mathbf{c}^{|K|}$). Now suppose that the outcome distributions $\boldsymbol{\pi}_a^k$ are unknown, except for the following (strict) first-order stochastic dominance relation: For any action a and outcome j, a type $k+1$ agent is more likely to obtain an outcome at least as good as j when taking action a as a type k agent. Here, agent $k+1$ can be said to be more efficient than agent k, in the sense that agent $k+1$ is more productive under each possible action. This ordering between agent types can be captured by the following parameter set:

$$\Pi = \left\{ \boldsymbol{\pi} \in Q_\pi \middle| \boldsymbol{\pi} \geq \mathbf{0}, \sum_{j \in J} \pi_{aj}^k = 1 \text{ for } a \in A, k \in K \right\}, \tag{A.5}$$

where

$$Q_\pi = \left\{ \boldsymbol{\pi} \middle| \sum_{h=j}^{d} \pi_{ah}^{k} + \epsilon \leq \sum_{h=j}^{d} \pi_{ah}^{k+1}, \text{ for } a \in \{1, 2, \ldots, m-1\}, j \in J, k \in \{1, 2, \ldots, |K|-1\} \right\}, \tag{A.6}$$

where $\epsilon > 0$ is a small constant. The inequalities given in (A.6) restrict the parameter set so that the outcome distribution is stochastically dominated when taken by a higher type (i.e., more efficient) agent. For an example where this kind of ordering may arise, consider an employee bonus program where the action set A corresponds to employee effort levels, and employees are one of two types: inexperienced ($k = 1$) or experienced ($k = 2$). Here, the outcome distribution $\boldsymbol{\pi}_a^k$ may be interpreted as a type k agent's productivity under action a. If the productivity of both agent types is unknown, but there is other evidence to suggest that experienced agents are more productive than inexperienced agents (all else equal), then we might use a parameter set like the one given in (A.5) to simultaneously estimate $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$, which can be interpreted as the agents' production functions.

Cost efficiency.

Agents may also be heterogeneous with respect to action costs. Naturally, we can model heterogeneity in agent costs by assuming all agents have the same outcome distributions ($\boldsymbol{\pi}^1 = \boldsymbol{\pi}^2 = \ldots = \boldsymbol{\pi}^{|K|}$), and assuming that action costs decrease in agent type: $c_a^1 \geq c_a^2 \geq \ldots \geq c_a^{|K|}$, for all $a \in A$. A more general and richer model for heterogeneity in action costs is the following:

$$Q_\pi = \left\{ \boldsymbol{\pi} \middle| \sum_{h=j}^{d} \pi_{ah}^{k+1} \leq \sum_{h=j}^{d} \pi_{bh}^{k} \text{ only if } c_b^{k+1} \geq c_a^k \right\}. \tag{A.7}$$

The parameter set in (A.7) implies that high-type agents are more cost efficient than low-type agents in the following sense: an action b taken by a low-type agent can only (weakly) stochastically dominate (i.e. be more productive than) an action a of a high-type agent if it is more costly for the low-type agent. In other words, for the low-type agent, there exists no action that is both less costly and more productive than any action of a high-type agent.

Limiting model complexity.

One potential challenge associated with our approach to incorporating agent heterogeneity is that it increases the number of parameters to be estimated; for instance, in the two examples above, the number of free parameters in the model is on the order of $m \cdot d \cdot |K|$. As with any model fitting problems, if the amount of available data is limited relative to the number of free parameters, there is a risk of the model overfitting the data, which leads to poor out-of-sample performance. A typical approach to preventing overfitting is to limit or penalize model complexity (i.e., through regularization). Within our framework, model complexity can be limited by imposing additional constraints on the parameter set Π. In particular, consider the following parameter set:

$$\Pi_{\text{reg}} = \left\{ \boldsymbol{\pi} \in \Pi \;\middle|\; \sum_{(a,b) \in A \times A} \sum_{(k,k') \in K \times K} \mathbf{1}\{\boldsymbol{\pi}_a^k \neq \boldsymbol{\pi}_b^{k'}\} \leq \ell \right\}, \tag{A.8}$$

where ℓ is an integer and $\ell \leq m \cdot |K|$. The set Π_{reg} may be interpreted as a regularized counterpart to Π. In words, Π_{reg} permits at most ℓ unique distributions to be used by the estimator PA-H in the construction of the matrices $\boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \ldots, \boldsymbol{\pi}^{|K|}$. Note that this regularization constraint can easily be imposed in the optimization formulation for PA-H (given in A.3.3 below) through constraints on the binary assignment variables $\mathbf{w}$.

A.3.2 Statistical consistency

Next, we show that the consistent estimation is possible under agent heterogeneity as well. Similar to (2.5), it will be helpful to define

$$R_a^k(\boldsymbol{\pi}) = \left\{ \mathbf{r} \in R \middle| a \in \operatorname*{argmax}_{a \in A} \sum_{j \in J} \pi_{aj}^k r_j - c_a^k \right\}.$$

Next, we present two assumptions that parallel Assumptions 1 and 2.

Assumption 7 (Data). *The data* $(\mathbf{r}^i, \xi^i, \theta^i)$ *are independent samples of random variables* $(\mathbf{r}, \xi, \theta)$*, where (i)* $(\mathbf{r}, \xi, \theta)$ *is jointly distributed with support* $R \times J \times K$*, (ii)* $\mathbf{r}$ *has continuous marginal density function* $f(\mathbf{r})$*, (iii)* ξ *has conditional mass function* $\pi_{aj}^{0k} = \Pr(\xi = j | \mathbf{r} \in R_a^k(\boldsymbol{\pi}))$*, and (iv) for any* $\boldsymbol{\pi} \in \Pi$*,* $\Pr(\mathbf{r} \in R_a^k(\boldsymbol{\pi})) > 0$ *for all* $a \in A$*,* $j \in J$*, and* $k \in K$*.*

Assumption 8 (Identifiability). *For every* $\boldsymbol{\pi} \in \Pi$ *such that* $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$*, there exists an* (a, j, k) *such that*

$$\pi_{aj}^k \neq \sum_{b \in A} \pi_{bj}^{0k} \cdot \Pr(\mathbf{r} \in R_b(\boldsymbol{\pi}^{0k}) | \mathbf{r} \in R_a(\boldsymbol{\pi}^k)).$$

We can now present a corollary to Theorem 1, which establishes that PA-H provides consistent estimates.

Corollary 1. *Let Assumption 7 hold. Then* $\hat{\boldsymbol{\pi}}_n \longrightarrow \boldsymbol{\pi}_0$ *if and only if Assumption 8 holds.*

Corollary 1 confirms that the estimates produced by PA-H converge to the true model parameters. Next, we show how the integer programming formulation PA-D and the statistical column generation algorithm given in §2.4.3 can be extended to incorporate agent heterogeneity.

A.3.3 Optimization, solution algorithm, and numerical examples

Similar to the single type estimator PA, we can formulate a proxy estimator for PA-H, represent this proxy estimator exactly as a mixed-integer formulation, and then obtain an integer programming formulation of a restricted estimator that limits each distribution $\boldsymbol{\pi}_a^k$ to a set of candidate distributions V. Because this development follows in a parallel manner to §2.3, we skip directly to the formulation of the restricted estimator.

Let $\mathbf{w}$, $\mathbf{x}$ and $\boldsymbol{\phi}$ be binary variables with the following interpretations, parallel to §2.4.1: $w_{as}^k = 1$ if the candidate distribution $\mathbf{v}_s$ is assigned to distribution $\boldsymbol{\pi}_a^k$, $x_s^i = 1$ if the action that candidate distribution $\mathbf{v}_s$ is assigned to is optimal under contract $\mathbf{r}^i$, and $\phi_{as}^i = 1$ if $\mathbf{v}_s$ is assigned to distribution $\boldsymbol{\pi}_a^k$ *and* action a is optimal under $\mathbf{r}^i$ and $\boldsymbol{\pi}^k$. Let $\mathbf{z}^k \in \mathbb{R}_+^{d \times S}$ be auxiliary variables. Then, analogous to PA-D, we can formulate the restricted estimator as the following optimization problem:

$$\underset{\boldsymbol{\pi},\mathbf{w},\mathbf{x},\mathbf{z},\boldsymbol{\phi}}{\text{minimize}} \quad \sum_{k\in K}\sum_{s\in S}\sum_{j\in J} z_{sj}^k \tag{A.9a}$$

$$\text{subject to} \quad z_{sj}^k \geq \frac{1}{n}\sum_{i\in I^k}(\xi_j^i - v_{sj})x_s^i, \quad j\in J, s\in S, k\in K, \tag{A.9b}$$

$$z_{sj}^k \geq \frac{1}{n}\sum_{i\in I^k}(v_{sj} - \xi_j^i)x_s^i, \quad j\in J, s\in S, k\in K, \tag{A.9c}$$

$$\sum_{b\in A}\sum_{s\in S}\left(\sum_{j\in J} v_{sj}r_j^i - c_b^k\right)\phi_{bs}^i \geq \left(\sum_{j\in J} v_{s'j}r_j^i - c_a^k\right)w_{as'}^k, \quad i\in I^k, a\in A, s'\in S, k\in K, \tag{A.9d}$$

(PA-DH)

$$\sum_{s\in S} w_{as}^k = 1, \quad a\in A, k\in K, \tag{A.9e}$$

$$\sum_{a\in A}\sum_{s\in S}\phi_{as}^i = 1, \quad i\in I, \tag{A.9f}$$

$$x_s^i = \sum_{a\in A}\phi_{as}^i, \quad i\in I, s\in S, \tag{A.9g}$$

$$\phi_{as}^i \leq w_{as}^k, \quad i\in I^k, a\in A, s\in S, k\in K, \tag{A.9h}$$

$$\pi_{aj}^k = \sum_{s\in S} w_{as}^k v_{sj}, \quad a\in A, j\in J, k\in K, \tag{A.9i}$$

$$x_s^i \in \{0,1\}, \quad i\in I, s\in S, \tag{A.9j}$$

$$w_{as}^k \in \{0,1\}, \quad a\in A, s\in S, k\in K, \tag{A.9k}$$

$$\phi_{as}^i \in \{0,1\}, \quad i\in I, a\in A, s\in S, \tag{A.9l}$$

$$\boldsymbol{\pi} \in Q_\pi. \tag{A.9m}$$

The candidate distributions in the formulation above can be constructed in an analogous way to Algorithm 1. The main difference is that we first segment the data according to agent type, which is the key step in obtaining "good" candidate distributions. Recall that the intuition behind

the sample-based construction of candidate distributions (given in Algorithm 1) is to use the data to approximate the true outcome distributions, $\boldsymbol{\pi}_a^0$, $a \in A$. Because each cluster in Algorithm 1 is constructed to contain *similar* contracts, we anticipate that for some clusters, all contracts that are contained within it induce the same (hidden) action from the agent. If, for a given cluster, all contracts lead to the same hidden action, then the empirical mass function over outcomes (i.e. the candidate distribution) should be informative about the outcome distribution $\boldsymbol{\pi}_a^0$ for some unknown $a \in A$. In the case where agents are heterogeneous, segmenting the data by agent types before constructing the candidate distributions helps to preserve this information. A summary is given in Algorithm 5.

Algorithm 5: Construction of candidate distributions with agent heterogeneity

Input: Data $(\mathbf{r}^i, \xi^i, \theta^i), i \in I$, parameter, $\rho > 0$.
1. For each $k \in K$, randomly sample a subset S^k from I^k.
2. **for** $s \in S^k$, $k \in K$:
   $B_s = \{\mathbf{r} \in R | \|\mathbf{r}^s - \mathbf{r}\|_2 \leq \rho\}$,
   $I_s = \{i \in I^k | \mathbf{r}^i \in B_s\}$.
   for $j \in J$:
   $v_{sj} = \frac{1}{|I_s|} \sum_{i \in I_s} y_j^i$.

Output: Candidate distributions $V = \{\mathbf{v}_s \text{ for } s \in S^k, k \in K\}$.

Next, we discuss how to use the statistical column generation algorithm to return a sufficiently "representative" subset of the candidate distributions in the presence of agent heterogeneity. Recall that the statistical column generation algorithm terminates when there is no candidate distribution in the set of omitted distributions V^- that is statistically different from every distribution in V^+. We use the same argument to build the corresponding algorithm in the presence of agent heterogeneity. We first define a vector $\boldsymbol{\psi}_s^k \in Z_+^d$, where the j^{th} entry is the frequency of outcome j in the candidate distribution $\mathbf{v}_s$ for $s \in S^k$, obtained in Algorithm 5. Then using the test functions introduced in §2.4.3, we identify a subset of candidate distributions V^+ and solve PA-HD over V^+

instead of V. Algorithm 6 provides an overview. Note that when there is no additional structure that links the agent types, then Algorithm 5 and 4 is equivalent to running Algorithm 1 and 2 for each agent type separately.

Algorithm 6: Statistical column generation (PA-DH+)

Input: Data $(\mathbf{r}^i, \xi^i, \theta^i)$, $i \in I$, candidate distributions V produced by Algorithm 5, significance level $\alpha > 0$.

Initialize: Set $t = 0$. Select any $s \in S^k$, $k \in K$. Set $S^+ = \{(s,k)\}$ and $S^- = S \setminus \{(s,k)\}$.

1. Let $(s^*, k^*) = \text{argmax}_{(s,k)\in S^-} \inf_{(s',k')\in S^+} \{H_\alpha(\boldsymbol{\psi}_s^k, \boldsymbol{\psi}_{s'}^{k'})\}$.

if $\inf_{(s',k')\in S^+} \left\{H_\alpha(\boldsymbol{\psi}_{s^*}^{k^*}, \boldsymbol{\psi}_{s'}^{k'})\right\} \leq 0$ or $S^- = \emptyset$,

Solve PA-HD(S^+) and obtain solution $\boldsymbol{\pi}_n^+$, set $T = t$, and **terminate**.

else Update $t \leftarrow t+1$, $S^+ \leftarrow \{S^+, (s^*, k^*)\}$, and $S^- \leftarrow S^- \setminus \{(s^*, k^*)\}$. Return to Step 1.

Output: Parameter estimate $\boldsymbol{\pi}_n^+$, iteration count T.

We conclude this section by providing a simple numerical example to investigate how incorporating *a priori* information about the relative efficiency of agents affects estimation error, compared to naively applying our base estimator to each agent type separately. We consider two agent types, L and H, which have true parameters $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$, respectively. Suppose a type H agent is known to be more outcome-efficient than a type L agent (§A.3.1). Then the unknown parameters $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$ must satisfy the following inequalities, where $\epsilon > 0$ is a small constant:

$$\sum_{h=j}^{d} \pi_{ah}^{0L} + \epsilon \leq \sum_{h=j}^{d} \pi_{ah}^{0H}, \text{ for } a \in \{1, 2, \ldots, m-1\},\ j \in J. \tag{A.10}$$

Recall that m and d denote the number of actions and outcomes, respectively. We consider two problem sizes, given by $(m,d) \in \{(2,2),(4,5)\}$. For each of the two problem sizes, we consider three sample sizes, given by $n \in \{100, 500, 1000\}$. Then for each combination (m,d,n), we randomly generate $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$ from Π given by (A.5), where Q_π is given by (A.6). We use the same data generation procedure described in §2.4.4.

			Without outcome efficiency			With outcome efficiency		
m	d	n	Type L	Type H	Mean	Type L	Type H	Mean
2	2	100	0.135	0.112	0.123	0.104	0.080	0.092
2	2	500	0.094	0.095	0.095	0.069	0.091	0.080
2	2	1000	0.077	0.060	0.069	0.084	0.085	0.084
4	5	100	0.058	0.052	0.055	0.054	0.043	0.049
4	5	500	0.054	0.053	0.054	0.043	0.040	0.041
4	5	1000	0.042	0.044	0.043	0.052	0.041	0.046

Table A.4: Normalized estimation error with and without outcome efficiency constraints, averaged over 10 trials.

Table A.4 summarizes the estimation errors over 10 trials for both agent types, with and without including (A.10) in the parameter set Q_π during estimation. Observe that the mean estimation error tends to decrease when the outcome efficiency information represented by the inequalities (A.10) is included, when $n = 100$ and $n = 500$. Conversely, for $n = 1000$, the estimation error *increases* when the outcome efficiency information is incorporated. The intuition for this result is as follows. Note that including the constraint (A.10) in Q_π can be interpreted as restricting the search space for the estimator LPA-DH. When n is small, the candidate distributions are poor approximations of the rows of $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$, and this restriction of the search space steers the estimator LPA-DH toward candidate distributions that approximate $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$ well. However, when the sample size is large, the candidate distributions are good approximations of the rows of $\boldsymbol{\pi}^{0L}$ and $\boldsymbol{\pi}^{0H}$, and the constraint (A.10) restricts the search space for the estimator without providing additional information, which has the net effect of causing errors to increase.

A.4 A Dynamic Principal-Agent Model with Hidden Actions

Our focus in this paper has been on estimating the agent model in an offline setting, where all data on historical contracts and outcomes is available at the outset, and the contract data $\mathbf{r}^i$, $i \in I$

is given exogenously. In this section, we consider an online counterpart to our model, where the principal can select the contracts to be offered to the agent in a dynamic manner. We note here that while extending our principal-agent framework to a dynamic setting raises new and important theoretical questions, a complete treatment of the dynamic setting is beyond the scope of this paper. Therefore, our focus will be to present results that parallel the offline setting, namely, presenting a complete solution algorithm for the dynamic contracting problem, and proving consistency of the corresponding estimates.

The remainder of this section is organized as follows. In §A.4.1, we briefly discuss related literature. In §A.4.2, we formulate a dynamic variant of our principal-agent model. In §A.4.3, we present an "ϵ-greedy" algorithm for the dynamic contracting problem, which involves iteratively solving an integer program and a sequence of linear programs. In §A.4.4, we present two consistency results related to the ϵ-greedy algorithm, and illustrate its performance with a simple numerical example. Proofs are contained in §A.5.4.

A.4.1 Related literature

There is an extensive literature on dynamic principal-agent models with hidden actions; for examples of foundational work, see Radner (1981), Rogerson (1985), Spear and Srivastava (1987), and Abreu et al. (1990). This line of research has typically focused on characterizing the principal's optimal decisions in an environment where parameters of the agent model are known, including the stochastic dependence of outcomes on agent actions.

The dynamic setting we consider is closer to the *multi-armed bandit problem*, which is a broad modeling framework for dynamic decision-making problems (see Slivkins (2019) for a recent review of bandit algorithms). A general setup for the multi-armed bandit problem with stochastic rewards is as follows. In each of T rounds, a utility-maximizing decision maker chooses from a set of actions

("arms"). The chosen arm generates a reward for the decision-maker, where the reward is an i.i.d. sample from an unknown distribution that depends on the arm. Because the reward distribution for each arm is unknown, the decision-maker faces what is often called an *exploration-exploitation* trade-off: they must balance exploiting "good" arms (which have been observed to produce high rewards in previous rounds) with exploring the full set of arms (which will improve knowledge about each arm's reward distribution).

Note that our paper focuses on estimating the distribution over outcomes induced by each possible agent action. For this reason, it is natural to use the multi-armed bandit framework to formulate a dynamic variant of our principal-agent model, where each arm represents a contract, and the selected contract generates a random reward for the principal (via the agent's hidden action). Further, unlike the offline setting, here we assume that the principal has preferences over the outcome set J. This setting is similar to work by Ho et al. (2016), who also address a dynamic principal-agent problem within a multi-armed bandit framework. Similar to our paper, the authors consider a principal-agent problem where a contract is a mapping of agent payments to outcomes, and outcomes depend on an unobservable agent action. A key distinction between Ho et al. (2016) and our work is that they assume the distribution over agent types is unknown, whereas we assume the dependence of outcomes on agent actions is the unknown parameter.

A.4.2 Model

The principal-agent interaction proceeds over T time periods. In period t, the principal selects and offers to the agent a contract $\mathbf{r}^t \in R$. As in the offline setting, the agent selects the action that maximizes their expected utility under the selected contract. As before, the agent's action remains hidden to the principal, who instead observes an outcome $\xi^t \in J$ in each period. If the agent takes action a, then $\xi^t = j$ with probability π^0_{aj}, where $\boldsymbol{\pi}^0$ is unknown to the principal. Upon observing

the outcome ξ^t, the principal selects a new contract $\mathbf{r}^{t+1}$, which initiates the next period.

Note that principal's preferences over the outcomes J are irrelevant for estimation in the offline setting, because the contract data is already fixed. However, because in the dynamic setting the principal now chooses the contracts $\mathbf{r}^t$, they must also balance estimation of $\boldsymbol{\pi}^0$ with maximizing their own utility over the T rounds. Let $\zeta_j(\mathbf{r})$ be the principal's utility under outcome $\{\xi^t = j\}$, which may include the payment r_j to the agent. Recall that $a(\mathbf{r})$ is the optimal action of the agent under $\mathbf{r}$ and $\boldsymbol{\pi}^0$; the principal's expected payoff under a contract $\mathbf{r}$ is then

$$U(\mathbf{r}) = \sum_{j \in J} \pi^0_{a(\mathbf{r}),j} \zeta_j(\mathbf{r}). \tag{A.11}$$

The principal's dynamic contracting problem is to select the sequence of contracts $\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^T$ that maximizes their expected payoff over T periods:

$$\max_{\mathbf{r}^t \in R} \sum_{t=1}^{T} U(\mathbf{r}^t). \tag{A.12}$$

We shall call any $\mathbf{r}^*$ that satisfies $\mathbf{r}^* \in \text{argmax}_{\mathbf{r} \in R} U(\mathbf{r})$ an *optimal* contract. Note that the optimal solution to (A.12) is to simply let $\mathbf{r}^t = \mathbf{r}^*$ for all $t \geq 1$. However, solving (A.12) is challenging because the principal's utility function $U(\mathbf{r})$ – and therefore the optimal contract $\mathbf{r}^*$ – depends on the unknown parameter $\boldsymbol{\pi}^0$. Therefore, the principal must trade-off learning $U(\mathbf{r})$ (by estimating $\boldsymbol{\pi}^0$) with maximizing $U(\mathbf{r})$ (by selecting high-utility contracts).

A.4.3 Algorithm overview

We now present a solution algorithm for the principal's dynamic contracting problem (A.12). Our approach is similar to ϵ-greedy algorithms found in the multi-armed bandit literature, which are intuitively simple and have been observed to perform well empirically (Kuleshov and Precup, 2014).

Within an ϵ-greedy framework, the decision-maker chooses an exploratory action in period t with probability ϵ^t, and chooses a utility-maximizing action with probability $1 - \epsilon^t$. Here, "greediness" refers to naively maximizing the decision maker's utility using the incumbent parameter estimates. In our setting, exploration means exploring the contract set R, which is necessary for learning the parameter $\boldsymbol{\pi}^0$; exploitation means selecting the contract in R that maximizes the principal's expected single-period utility $U(\mathbf{r})$, based on the current estimate of $\boldsymbol{\pi}^0$.

Before presenting the steps of our ϵ-greedy algorithm, we first define two new optimization problems: one for each of the exploration and exploitation steps of the algorithm.

Exploration step.

Note that the decision set to be explored is the continuous set R. To make exploration of R tractable, we first randomly sample a set of contracts $\mathbf{r}_s$, $s \in S$ from the contract set R, which remain fixed for the entire algorithm. Here, each $\mathbf{r}_s$, $s \in S$ is analogous to one "arm" in the multi-armed bandit setting. Let n_s^t be the number of times $\mathbf{r}_s$ has been selected after the t^{th} round. Then we can construct the candidate distribution associated with $\mathbf{r}^s$ as

$$\mathbf{v}_s^t = \frac{1}{n_s^t} \sum_{i=1}^{t} \mathbf{1}\{\xi^i = j\}.$$

We assume that the agent's true optimal action $a(\mathbf{r}^s)$ is unique under each $\mathbf{r}^s$ (note that this is almost surely the case if the initial arms $\mathbf{r}^s$, $s \in S$ are selected by randomly sampling from R according to a continuous distribution). Each $\mathbf{v}_s^t$ can then be interpreted as an empirical distribution constructed by sampling n_s^t times from the distribution $\pi^0_{a(\mathbf{r}^s)}$.

Next, define a set of error variables $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{|S|}$, and let w_{as} be a binary variable equal to 1

if $\mathbf{v}_s$ is assigned to $a \in A$. We now define the following optimization problem:

$$\underset{w,\varepsilon}{\text{minimize}} \quad \sum_{s\in S} |\varepsilon_s| \tag{A.13a}$$

$$\text{subject to} \quad \sum_{a\in A}\left(\sum_{j\in J} v_{sj}^t r_{sj} - c_a\right) w_{as} + \varepsilon_s \geq \sum_{a\in A}\left(\sum_{j\in J} v_{s'j}^t r_{sj} - c_a\right) w_{as'}, \quad s \in S, s' \in S, \tag{A.13b}$$

$$\sum_{a\in A} w_{as} = 1, \quad s \in S, \tag{A.13c}$$

(PA-T)

$$\sum_{s\in S} w_{as} \geq 1, \quad a \in A, \tag{A.13d}$$

$$w_{as} \in \{0,1\}, \quad a \in A, s \in S. \tag{A.13e}$$

$$\varepsilon_s \geq 0, \quad s \in S. \tag{A.13f}$$

Intuitively, formulation PA-T assigns each contract $\mathbf{r}^s$ to the action $a \in A$ that is believed to be optimal under $\mathbf{r}^s$ and the true model $\boldsymbol{\pi}^0$. Constraint (A.13c) ensures every contract $\mathbf{r}^s$ is assigned to exactly one action, and (A.13d) ensures every action has at least one contract assigned to it. The key constraint of this formulation is (A.13b): Given a contract $\mathbf{r}^s$, constraint (A.13b) forces the action associated with $\mathbf{r}^s$ to yield at least as large of a utility for the agent as the action associated with $\mathbf{r}^{s'}$, for every $s' \in S$. Note that because $\mathbf{v}_s^t$ is an approximation of $\boldsymbol{\pi}_a^0$ for some $a \in A$, constraint (A.13b) may not hold when $\varepsilon_s = 0$, even if the binary decision variable $\mathbf{w}$ correctly assigns each $\mathbf{r}^s$ to the agent's true optimal action. Therefore, ε_s is required to serve as a slack variable that maintains feasibility of PA-T. Intuitively, ε_s can be interpreted as measuring the sub-optimality of the assignment encoded by $\mathbf{w}$.

Let $(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t)$ be the optimal solution to PA-T in round t. Then we construct the estimate $\hat{\boldsymbol{\pi}}^t$ as

follows:

$$\hat{\pi}^t_{aj} = \frac{\sum_{s\in S} v^t_{sj} n^t_s \bar{w}^t_{as}}{\sum_{s\in S} n^t_s \bar{w}^t_{as}}, \quad \text{for } a \in A, j \in J.$$

Intuitively, $\hat{\boldsymbol{\pi}}^t_a$ is the average of all candidate distributions assigned to action a, weighted by their sample sizes n^t_s.

Exploitation step.

The exploitation step consists of solving the optimal contracting problem $\mathsf{OC}(\hat{\boldsymbol{\pi}}^t)$, described in §A.1 of the electronic companion, under the incumbent estimate $\hat{\boldsymbol{\pi}}^t$.

Algorithm summary.

Having defined the optimization models PA-C and OC, we can now summarize the steps of the ϵ-greedy algorithm (Algorithm 7). At each iteration, an exploration action is taken with probability ϵ^t, by randomly selecting one of the contracts from $\{\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^{|S|}\}$ and updating the set of candidate distributions accordingly, and an exploitation action is taken with probability $1-\epsilon^t$, by selecting the optimal contract based on the current estimate, $\hat{\boldsymbol{\pi}}$.

ALGORITHM 7: ϵ-greedy algorithm for dynamic contracting

Input: Exploration parameters $\epsilon^1, \epsilon^2, \ldots, \epsilon^T$.
1. **Initialize**: Randomly sample S contracts $\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^S$ from set R.
   Initialize $n^t_s = 1$ for $s \in S$ and $t = 1$.
2. **for** $t = 1, 2, \ldots, T$:
   with probability ϵ^t:
   Set $\hat{\mathbf{r}}^t = \hat{\mathbf{r}}^{t-1}$.
   Randomly select $\bar{s} \in S$. Select contract $\mathbf{r}^{\bar{s}}$ and observe outcome $\bar{\xi}$.
   Set $s^t = \bar{s}$ and $\xi^t = \bar{\xi}$.
   Update $n^t_{\bar{s}} \leftarrow n^t_{\bar{s}} + 1$ and $v^t_{\bar{s}j} = \frac{1}{n^t_{\bar{s}}} \sum_{i=1}^t \mathbf{1}\{\xi^i = j, s^i = \bar{s}\}$ for $j \in J$.
   with probability $1-\epsilon^t$:
   Solve PA-T and obtain $\bar{\mathbf{w}}^t$. For each $a \in A$, $j \in J$, set $\hat{\pi}^t_{aj} = \frac{\sum_{s\in S} v^t_{sj} n^t_s \bar{w}^t_{as}}{\sum_{s\in S} n^t_s \bar{w}^t_{as}}$.
   Solve $\mathsf{OC}(\hat{\boldsymbol{\pi}}^t)$ and obtain $\hat{\mathbf{r}}^t$.
Output: Estimates $\hat{\boldsymbol{\pi}}^T$ and $\hat{\mathbf{r}}^T$.

A.4.4 Consistency, numerical examples, and discussion

We now present the main theoretical result of this section. First, consider the following assumption:

Assumption 9. *For each $s \in S$, there exists $s' \in S$ such that*

$$\sum_{j \in J} (\pi^0_{a(\mathbf{r}^{s'}),j} - \pi^0_{a(\mathbf{r}^s),j}) r^s_j > \bar{c} - \underline{c}, \tag{A.14}$$

where $\bar{c} = \sup_{a \in A}\{c_a\}$ and $\underline{c} = \inf_{a \in A}\{c_a\}$.

Assumption 9 is an identifiability condition that ensures precise inference of $\boldsymbol{\pi}^0$, similar to Assumption 2 in the offline setting. Loosely speaking, this condition ensures that if there exists a solution $\bar{\mathbf{w}}$ to PA-T that assigns a candidate distribution v_s to the wrong action, then the assignments encoded in $\bar{w}$ will eventually (as more data is collected) violate the optimality conditions in (A.13b). This has the effect of guaranteeing that the solution to PA-T identifies the correct action for each candidate distribution v_s in the limit. Next, Proposition 5 states that the condition in 9 is sufficient for Algorithm 7 to uncover the true values of $\boldsymbol{\pi}^0$.

Proposition 5. *Let Assumption 9 hold. Then the estimate $\hat{\boldsymbol{\pi}}^T$ produced by Algorithm 7 is consistent:*

$$\hat{\boldsymbol{\pi}}^T \longrightarrow \boldsymbol{\pi}^0.$$

Note that Assumption 9 is stronger than the analogous condition in the offline setting, Assumption 8). The reason for this difference is as follows: In the online setting, we restrict attention to the subset of contracts $\{\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^{|S|}\}$ when estimating $\hat{\boldsymbol{\pi}}$. While focusing on this discrete subset of R improves tractability, it also reduces variation in the data compared to the offline setting, which reduces the information available to the estimator. Because the online setting has less variation in contract data, a stronger identifiability condition is required to ensure that learning the precise

values of the parameter $\boldsymbol{\pi}^0$ remains possible.

Corollary 2. *Let Assumption 9 hold. Then the contract $\hat{\mathbf{r}}^T$ produced by Algorithm 7 converges to a minimizer of the principal's per-period regret:*

$$\left| \frac{1}{T} \sum_{t=1}^{T} U(\hat{\mathbf{r}}^T) - \frac{1}{T} \sum_{t=1}^{T} U(\mathbf{r}^*) \right| \longrightarrow 0.$$

The intuition behind Corollary 2 is straightforward – because $\hat{\boldsymbol{\pi}}^T$ is a consistent estimate of $\boldsymbol{\pi}^0$, then solving $\mathsf{OC}(\hat{\boldsymbol{\pi}}^T)$ will eventually produce an optimal contract for the principal, as $T \longrightarrow \infty$.

We now illustrate Algorithm 7 through numerical examples. We consider two problem sizes, given by $(m, d) \in \{(2, 2), (5, 10)\}$. For each problem size, we construct the contract set as $R = [1, 2]^d$, the agent cost vector $\mathbf{c}$ as a random sample from $[0, 1]^m$. We specify the principal's utility as $\zeta_j(\mathbf{r}) = \bar{\zeta}_j - r_j$, where $\bar{\zeta}_j$ is a random sample from $[1, 10]^d$. We construct the true parameter $\boldsymbol{\pi}^0$ by letting $\boldsymbol{\pi}^0_a$ for each $a \in A$ be randomly generated from the $(d-1)$-dimension simplex. We test six parameterizations of the exploration probability ϵ^t. We first consider three "fixed" exploration schemes, where $\epsilon^t = \epsilon$, for each $\epsilon \in \{0.5, 0.9, 0.99\}$. We then consider three "variable" exploration schemes, where $\epsilon^t = \exp(-\lambda \cdot t)$, for each $\lambda \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. For each problem size and parameterization of ϵ^t, we run 10 trials of Algorithm 7 with $T = 1000$, where π^0, $\bar{\boldsymbol{\zeta}}$, and $\mathbf{c}$ are randomly constructed in each trial.

A summary of results is provided in Table A.5 ($m = 2, d = 2$) and Table A.6 ($m = 5, d = 10$). For each of the six exploration schemes, we report the average estimation error and average per-period regret at $t = 100$, $t = 500$ and $t = 1000$. Among the fixed exploration schemes, aggressive exploitation ($\epsilon^t = 0.5$) attains the lowest regret and highest estimation error, and aggressive exploration ($\epsilon^t = 0.99$) attain the highest regret and lowest estimation error. These results demonstrate the exploration-exploitation trade-off in our dynamic contracting problem. Comparing all six ex-

			$\epsilon^t = \epsilon$			$\epsilon^t = \exp(-\lambda \cdot t)$		
m	d	T	$\epsilon = 0.5$	$\epsilon = 0.9$	$\epsilon = 0.99$	$\lambda = 10^{-2}$	$\lambda = 10^{-3}$	$\lambda = 10^{-4}$
2	2	50	0.21	0.11	0.07	0.06	0.06	0.06
2	2	100	0.20	0.11	0.07	0.06	0.06	0.06
2	2	500	0.16	0.08	0.07	0.05	0.06	0.06
2	2	1000	0.13	0.08	0.07	0.06	0.06	0.06
2	2	50	0.12	0.16	0.14	0.10	0.14	0.16
2	2	100	0.09	0.16	0.15	0.09	0.14	0.16
2	2	500	0.08	0.14	0.16	0.03	0.11	0.16
2	2	1000	0.08	0.14	0.16	0.02	0.09	0.16

Table A.5: Estimation errors (top) and per-period regret (bottom) for six exploration schemes, averaged over 10 trials ($m = 2, d = 2$).

			$\epsilon^t = \epsilon$			$\epsilon^t = \exp(-\lambda \cdot t)$		
m	d	T	$\epsilon = 0.5$	$\epsilon = 0.9$	$\epsilon = 0.99$	$\lambda = 10^{-2}$	$\lambda = 10^{-3}$	$\lambda = 10^{-4}$
5	10	50	0.05	0.06	0.06	0.05	0.05	0.05
5	10	100	0.06	0.06	0.06	0.05	0.05	0.05
5	10	500	0.06	0.06	0.05	0.05	0.05	0.05
5	10	1000	0.06	0.06	0.05	0.05	0.05	0.05
5	10	50	0.78	0.82	0.87	0.78	0.84	0.85
5	10	100	0.77	0.83	0.87	0.74	0.83	0.84
5	10	500	0.78	0.82	0.85	0.58	0.79	0.83
5	10	1000	0.79	0.82	0.85	0.56	0.73	0.83

Table A.6: Estimation errors (top) and per-period regret (bottom) for six exploration schemes, averaged over 10 trials ($m = 5, d = 10$).

ploration schemes, the variable scheme with $\lambda = 10^{-2}$ appears to weakly dominate, by producing similar estimation errors as the other approaches, but with significantly lower regret. This suggests that $\lambda = 10^{-2}$ handles the exploration-exploitation trade-off the most efficiently. We also observe that the decrease in regret is significantly more pronounced for the variable scheme with $\lambda = 10^{-2}$ compared to the other approaches, because it exploits the most aggressively as t increases (e.g., by $t = 500$, this scheme exploits with probability 0.99.)

We conclude by highlighting directions for future work. First, the algorithm we presented in this section is based on ϵ-greedy approaches found in the multi-armed bandit literature. Other

well-known bandit algorithms may also be applicable in our setting, such as those based on upper confidence bounds or Thompson sampling (Slivkins, 2019). Second, while we have focused our analysis in this section on asymptotic results, the standard performance measure of a dynamic decision-making algorithm is a finite-sample bound on regret. It may be possible to obtain similar bounds in our dynamic contracting setting. However, doing so will require addressing two technical challenges of our setting that are not present in classical bandit problems: the continuous nature of the decision space R (instead of a discrete set of arms), and discontinuity in the principal's payoff in $\mathbf{r}$ (due to jumps in the agent's optimal action as $\mathbf{r}$ changes). Third, we have assumed throughout that the agent behaves myopically by optimizing their single-period payoff. It may also be the case that agents behave strategically, which would introduce new and challenging dynamics into the problem.

Note also that the performance of Algorithm 7 is sensitive to how R is specified. Because R is the exploration space, if the values in R are large compared to the principal's utility, then the algorithm may spend a large number of iterations exploring regions of R that are inefficient from the principal's perspective. This sensitivity to the contract set R occurs because the Algorithm 7 does not consider regret when selecting a contract to explore, which is common in ϵ-greedy approaches.

A.5 Proofs

Because the proof for Theorem 2 is long, we group all proofs into four subsections for ease of navigation: §A.5.1 contains the proofs for Theorem 1 and Proposition 1; §A.5.2 contains the proof of Theorem 2; §A.5.3 contains the proofs of Proposition 2, Theorem 3, and Theorem 4; and §A.5.4 contains the proofs for Sections A.3 and A.4.

A.5.1 Proofs of Theorem 1 and Proposition 1

Before proving Theorem 1, we prove two supporting results, given in Lemmas 3 and 4. Lemma 3 shows that the loss function $L_n(\boldsymbol{\pi})$ is lower semicontinuous. Lemma 4 shows that as $n \longrightarrow \infty$, the loss function $L_n(\boldsymbol{\pi})$ converges point-wise to a function $L(\boldsymbol{\pi})$, where $L(\boldsymbol{\pi})$ is uniquely minimized by the true parameter $\boldsymbol{\pi}^0$ if and only if Assumption 2 holds. Both of these results are used to prove the consistency result in Theorem 1.

Lemma 3. *$L_n(\boldsymbol{\pi})$ is lower semicontinuous in $\boldsymbol{\pi}$ on Π for all $n \geq 1$.*

Proof. Define $N(\delta) = \{\boldsymbol{\pi} \in \Pi | \|\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}\|_1 < \delta\}$. To show that $L_n(\boldsymbol{\pi})$ is lower semicontinuous at all $\bar{\boldsymbol{\pi}} \in \Pi$ and $n \geq 1$, it suffices to prove the following statement (Rockafellar and Wets, 2009): For every $\varepsilon > 0$, there exists $\delta > 0$ such that $L_n(\bar{\boldsymbol{\pi}}) - L_n(\boldsymbol{\pi}) < \varepsilon$ for all $\boldsymbol{\pi} \in N(\delta)$. The proof proceeds in two steps. First, we show that there exists $\bar{\delta} > 0$ such that for all $\delta \in (0, \bar{\delta})$ and $\boldsymbol{\pi} \in N(\delta)$, $A(\mathbf{r}^i, \boldsymbol{\pi}) \subseteq A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$ for all $i \in I$. Second, we show lower semicontinuity of $L_n(\boldsymbol{\pi})$ at $\bar{\boldsymbol{\pi}}$. Step 1. Fix $\bar{\boldsymbol{\pi}}$ and n. By way of contradiction, suppose that for all $\delta > 0$, there exists $\check{\boldsymbol{\pi}} \in N(\delta)$, $a \in A(\mathbf{r}^i, \check{\boldsymbol{\pi}})$ and $i \in I$ such that $a \notin A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$. Note that $a \notin A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$ and $a \in A(\mathbf{r}^i, \check{\boldsymbol{\pi}})$ implies there exists a $b \in A \setminus a$ such that the following two inequalities hold:

$$\sum_{j \in J} \bar{\pi}_{aj} r_j^i - c_a < \sum_{j \in J} \bar{\pi}_{bj} r_j^i - c_b \tag{A.15}$$

$$\sum_{j \in J} \check{\pi}_{aj} r_j^i - c_a \geq \sum_{j \in J} \check{\pi}_{bj} r_j^i - c_b. \tag{A.16}$$

Combining (A.15) and (A.16) yields

$$\sum_{j \in J} (\check{\pi}_{aj} - \bar{\pi}_{aj}) r_j^i - c_a > \sum_{j \in J} (\check{\pi}_{bj} - \bar{\pi}_{bj}) r_j^i - c_b.$$

Letting $\delta \longrightarrow 0$ implies $\|\check{\boldsymbol{\pi}} - \bar{\boldsymbol{\pi}}\|_1 \longrightarrow 0$ and thus $0 > 0$, a contradiction. It follows that there exists

$\bar{\delta} > 0$ such that for all $\delta \in (0, \bar{\delta})$ and $\boldsymbol{\pi} \in N(\delta)$, $A(\mathbf{r}^i, \boldsymbol{\pi}) \subseteq A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$ for all $i \in I$. Step 2. Fix $\varepsilon > 0$. By way of contradiction, suppose $L_n(\boldsymbol{\pi})$ is not lower semicontinuous; that is, for every $\delta > 0$, there exists $\check{\boldsymbol{\pi}}_\delta \in N(\delta)$ such that $L_n(\bar{\boldsymbol{\pi}}) - L_n(\check{\boldsymbol{\pi}}_\delta) > \varepsilon$. Next, let $(\mathbf{x}(\boldsymbol{\pi}), \boldsymbol{\omega}(\boldsymbol{\pi}))$ be an optimal solution to (2.4) under $\boldsymbol{\pi} = \check{\boldsymbol{\pi}}_\delta$. By Step 1, if $\delta \in (0, \bar{\delta})$, then $A(\mathbf{r}^i, \check{\boldsymbol{\pi}}_\delta) \subseteq A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$ for all $i \in I$. Because $A(\mathbf{r}^i, \check{\boldsymbol{\pi}}_\delta) \subseteq A(\mathbf{r}^i, \bar{\boldsymbol{\pi}})$ for all $i \in I$, it is feasible to set $\boldsymbol{\pi} = \bar{\boldsymbol{\pi}}$ and $(\mathbf{x}, \boldsymbol{\omega}) = (\mathbf{x}(\check{\boldsymbol{\pi}}_\delta), \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta))$ in (2.4). It follows that $L_n(\bar{\boldsymbol{\pi}}) = \|\bar{\boldsymbol{\pi}} - \boldsymbol{\omega}(\bar{\boldsymbol{\pi}})\|_1 \leq \|\bar{\boldsymbol{\pi}} - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1$. Therefore,

$$\varepsilon < L_n(\bar{\boldsymbol{\pi}}) - L_n(\check{\boldsymbol{\pi}}_\delta) \leq \|\bar{\boldsymbol{\pi}} - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1 - L_n(\check{\boldsymbol{\pi}}_\delta) = \|\bar{\boldsymbol{\pi}} - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1 - \|\check{\boldsymbol{\pi}} - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1.$$

Further, it is straightforward to verify that

$$\|\bar{\boldsymbol{\pi}} - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1 - \|\check{\boldsymbol{\pi}}_\delta - \boldsymbol{\omega}(\check{\boldsymbol{\pi}}_\delta)\|_1 \leq \|\bar{\boldsymbol{\pi}} - \check{\boldsymbol{\pi}}_\delta\|_1.$$

Thus, $\varepsilon \leq \|\bar{\boldsymbol{\pi}} - \check{\boldsymbol{\pi}}_\delta\|_1$. Letting $\delta \longrightarrow 0$ implies $\|\bar{\boldsymbol{\pi}} - \check{\boldsymbol{\pi}}_\delta\|_1 \longrightarrow 0$, which yields a contradiction. We conclude that for every $\varepsilon > 0$, there exists $\delta > 0$ such that $L_n(\bar{\boldsymbol{\pi}}) - L_n(\boldsymbol{\pi}) < \varepsilon$ for all $\boldsymbol{\pi} \in N(\delta)$, and thus $L_n(\boldsymbol{\pi})$ is lower semicontinuous. □

Lemma 4. *There exists $L(\boldsymbol{\pi}) : \Pi \longrightarrow \mathbb{R}$ such that (i) $L(\boldsymbol{\pi}) < \infty$ and $|L_n(\boldsymbol{\pi}) - L(\boldsymbol{\pi})| \longrightarrow 0$ for all $\boldsymbol{\pi} \in \Pi$, (ii) $L(\boldsymbol{\pi}^0) = 0$, and (iii) $L(\boldsymbol{\pi}^0) < L(\boldsymbol{\pi})$ for every $\boldsymbol{\pi} \in \Pi$ such that $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$ if and only if Assumption 2 holds.*

Proof. The proof proceeds in two steps. In Step 1, we show the following supporting result: For any (a, j),

$$\lim_{n\to\infty} \omega_{aj}(\boldsymbol{\pi}) = \frac{1}{\Pr(\mathbf{r} \in R_a(\boldsymbol{\pi}))} \sum_{b \in A} \pi^0_{bj} \cdot \Pr(\mathbf{r} \in R_a(\boldsymbol{\pi}), \mathbf{r} \in R_b(\boldsymbol{\pi}^0)), \tag{A.17}$$

where $R_a(\boldsymbol{\pi})$ is defined in (2.5). In Step 2, we prove there exists a function $L(\boldsymbol{\pi})$ such that

statements (i), (ii) and (iii) hold. Step 1. Pick any (a, j). Note that $x^i = a$ holds in (2.4) holds if and only if $\mathbf{r}^i \in R_a(\boldsymbol{\pi})$. For each (a, j), we can then write ω_{aj} as

$$\omega_{aj}(\boldsymbol{\pi}) = n \cdot \frac{1}{n} \cdot \omega_{aj}(\boldsymbol{\pi}) = \frac{n}{\sum_{i=1}^{n} \mathbb{I}\{\mathbf{r}^i \in R_a(\boldsymbol{\pi})\}} \left(\frac{1}{n} \sum_{i \in \{i | \mathbf{r}^i \in R_a(\boldsymbol{\pi})\}} y_j^i \right). \tag{A.18}$$

Consider the first term on the right-hand side of (A.18). By Assumption 1, letting $n \longrightarrow \infty$ yields

$$\frac{n}{\sum_{i=1}^{n} \mathbb{I}\{\mathbf{r}^i \in R_a(\boldsymbol{\pi})\}} = \frac{1}{\Pr(\mathbf{r} \in R_a(\boldsymbol{\pi}))}. \tag{A.19}$$

For the second term on the right-hand side of (A.18), we have

$$\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \sum_{i \in \{i | \mathbf{r}^i \in R_a(\boldsymbol{\pi})\}} y_j^i &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\xi^i = j, \mathbf{r}^i \in R_a(\boldsymbol{\pi})\}, \\
&= \Pr(\xi = j, \mathbf{r} \in R_a(\boldsymbol{\pi})), \\
&= \sum_{b \in A} \Pr(\xi = j, \mathbf{r} \in R_a(\boldsymbol{\pi}) | \mathbf{r} \in R_b(\boldsymbol{\pi}^0)) \cdot \Pr(\mathbf{r} \in R_b(\boldsymbol{\pi}^0)).
\end{aligned}$$

The first equality above follows because $y_j^i = \mathbb{I}\{\xi^i = j\}$, by definition. The second equality follows from the strong law of large numbers. The third equality follows from the law of total probability. Next, note that by Assumption 1, the events $\{\xi = j\}$ and $\{\mathbf{r} \in R_a(\boldsymbol{\pi})\}$ are conditionally independent given $\{\mathbf{r} \in R_b(\boldsymbol{\pi}^0)\}$. Therefore,

$$\begin{aligned}
\Pr(\xi = j, \mathbf{r} \in R_a(\boldsymbol{\pi}) | \mathbf{r} \in R_b(\boldsymbol{\pi}^0)) &= \Pr(\xi = j | \mathbf{r} \in R_b(\boldsymbol{\pi}^0)) \cdot \Pr(\mathbf{r} \in R_a(\boldsymbol{\pi}) | \mathbf{r} \in R_b(\boldsymbol{\pi}^0)), \\
&= \pi_{bj}^0 \cdot \Pr(\mathbf{r} \in R_a(\boldsymbol{\pi}) | \mathbf{r} \in R_b(\boldsymbol{\pi}^0)),
\end{aligned}$$

where the second equality follows by definition of $\boldsymbol{\pi}^0$. It follows that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in\{i|\mathbf{r}^i\in R_a(\boldsymbol{\pi})\}} y_j^i = \sum_{b\in A} \pi_{bj}^0 \cdot \Pr(\mathbf{r}\in R_a(\boldsymbol{\pi})|\mathbf{r}\in R_b(\boldsymbol{\pi}^0)) \cdot \Pr(\mathbf{r}\in R_b(\boldsymbol{\pi}^0)),$$

$$= \sum_{b\in A} \pi_{bj}^0 \cdot \Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}), \mathbf{r}\in R_b(\boldsymbol{\pi}^0)). \tag{A.20}$$

Next, combining (A.18), (A.19), and (A.20) yields

$$\lim_{n\to\infty} \omega_{aj}(\boldsymbol{\pi}) = \frac{1}{\Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}))} \sum_{b\in A} \pi_{bj}^0 \cdot \Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}), \mathbf{r}\in R_b(\boldsymbol{\pi}^0)),$$

as desired. Step 2. Define $L(\boldsymbol{\pi}) = \sum_{a\in A}\sum_{j\in J} |\pi_{aj} - \lim_{n\to\infty}\omega_{aj}(\boldsymbol{\pi})|$. We prove (i), (ii) and (iii) in order. (i). Note by continuity of the absolute value function, $\lim_{n\to\infty}|\pi_{aj} - \omega_{aj}(\boldsymbol{\pi})| = |\pi_{aj} - \lim_{n\to\infty}\omega_{aj}(\boldsymbol{\pi})|$ for all (a,j) and $\boldsymbol{\pi}\in\Pi$. It follows by definition of $L(\boldsymbol{\pi})$ that $|L_n(\boldsymbol{\pi}) - L(\boldsymbol{\pi})| \longrightarrow 0$. Further, $L(\boldsymbol{\pi}^0) < \infty$ follows by definition of $L(\boldsymbol{\pi})$ and because $\lim_{n\to\infty}\omega_{aj}(\boldsymbol{\pi}) < \infty$ by Step 1. (ii). By definition of $L(\boldsymbol{\pi})$, to show that $L(\boldsymbol{\pi}^0) = 0$, it suffices to show that $\pi_{aj}^0 = \lim_{n\to\infty}\omega_{aj}(\boldsymbol{\pi}^0)$ for all (a,j). Pick any (a,j), and note

$$\begin{aligned}
\pi_{aj}^0 &= \frac{1}{\Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0))} \cdot \pi_{aj}^0 \cdot \Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0)), \\
&= \frac{1}{\Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0))} \sum_{b\in A} \pi_{bj}^0 \cdot \Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0), \mathbf{r}\in R_b(\boldsymbol{\pi}^0)), \\
&= \lim_{n\to\infty} \omega_{aj}(\boldsymbol{\pi}^0),
\end{aligned}$$

where the first equality follows by multiplying and dividing by $\Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0))$, the second equality follows because $\Pr(\mathbf{r}\in R_a(\boldsymbol{\pi}^0), \mathbf{r}\in R_b(\boldsymbol{\pi}^0)) = 0$ for all $b \neq a$ by continuity of $f(\mathbf{r})$, and the third equality follows from Step 1. Therefore, $L(\boldsymbol{\pi}^0) = 0$. (iii). Next, we establish that $L(\boldsymbol{\pi}) > 0$ for all $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$ if and only if Assumption 2 holds. First, let Assumption 2 hold. Pick any $\bar{\boldsymbol{\pi}} \neq \boldsymbol{\pi}^0$, and

suppose by way of contradiction that $L(\bar{\boldsymbol{\pi}}) = 0$. It follows that for all (a, j),

$$\begin{aligned}
\bar{\pi}_{aj} &= \lim_{n\to\infty} \omega_{aj}(\bar{\boldsymbol{\pi}}), \\
&= \frac{1}{\Pr(\mathbf{r} \in R_a(\bar{\boldsymbol{\pi}}))} \sum_{b\in A} \pi^0_{bj} \cdot \Pr(\mathbf{r} \in R_a(\bar{\boldsymbol{\pi}}), \mathbf{r} \in R_b(\boldsymbol{\pi}^0)), \\
&= \sum_{b\in A} \pi^0_{bj} \cdot \Pr(\mathbf{r} \in R_b(\boldsymbol{\pi}^0) | \mathbf{r} \in R_a(\bar{\boldsymbol{\pi}})),
\end{aligned}$$

where the first line follows because $L(\bar{\boldsymbol{\pi}}) = 0$, the second line follows from Step 1, and the third line follows by the probability chain rule. However, by Assumption 2, there exists an (a, j) such that $\bar{\pi}_{aj} \neq \sum_{b\in A} \pi^0_{bj} \cdot \Pr(\mathbf{r} \in R_b(\boldsymbol{\pi}^0) | \mathbf{r} \in R_a(\bar{\boldsymbol{\pi}}))$, which yields a contradiction. Therefore, $L(\boldsymbol{\pi}) > 0$ for all $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$. Conversely, if Assumption 2 does not hold, then by parallel argument to the above, there exists $\check{\boldsymbol{\pi}} \in \Pi$ where $\check{\boldsymbol{\pi}} \neq \boldsymbol{\pi}^0$ such that $\check{\pi}_{aj} = \lim_{n\to\infty}(\check{\boldsymbol{\pi}})$ for all (a, j). By definition of $L(\boldsymbol{\pi})$, it follows that $L(\boldsymbol{\pi}^0) = L(\check{\boldsymbol{\pi}}) = 0$. □

Proof of Theorem 1. The proof proceeds in two steps. First, we show $|L_n(\hat{\boldsymbol{\pi}}_n) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$. Second, we show $\text{plim}_{n\to\infty} \hat{\boldsymbol{\pi}}_n = \boldsymbol{\pi}^0$ if and only if Assumption 2 holds. Step 1. Because $L_n(\boldsymbol{\pi}^0) \longrightarrow L(\boldsymbol{\pi}^0)$ by Lemma 4(i) and $L(\boldsymbol{\pi}^0) = 0$ by Lemma 4(ii), we have $L_n(\boldsymbol{\pi}^0) \longrightarrow 0$. Next, note

$$0 \leq L_n(\hat{\boldsymbol{\pi}}_n) \leq L_n(\boldsymbol{\pi}^0),$$

where the first and second inequalities follow by definition of $L_n(\boldsymbol{\pi})$ and $\hat{\boldsymbol{\pi}}_n$, respectively. It follows that $L_n(\hat{\boldsymbol{\pi}}_n) \longrightarrow 0$. Therefore, $|L_n(\hat{\boldsymbol{\pi}}_n) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$. Step 2. Because $L_n(\boldsymbol{\pi})$ is lower semicontinuous by Lemma 3, $|L_n(\hat{\boldsymbol{\pi}}_n) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$ by Step 1, and Π is compact, by Theorem 5.14 of Van der Vaart (2000), $\text{plim}_{n\to\infty} \hat{\boldsymbol{\pi}}_n \in \text{argmin}_{\boldsymbol{\pi}\in\Pi} L(\boldsymbol{\pi})$. Next, suppose Assumption 2 holds. Then by Lemma 4, $\text{argmin}_{\boldsymbol{\pi}\in\Pi} L(\boldsymbol{\pi}) = \boldsymbol{\pi}^0$, which implies $\text{plim}_{n\to\infty} \hat{\boldsymbol{\pi}}_n = \boldsymbol{\pi}^0$. If Assumption 2 does

not hold, then by Lemma 4(iii), there exists $\tilde{\boldsymbol{\pi}} \neq \boldsymbol{\pi}^0$ such that $L(\tilde{\boldsymbol{\pi}}) = L(\boldsymbol{\pi}^0) = 0$, in which case $\text{plim}_{n\to\infty}\hat{\boldsymbol{\pi}}_n \neq \boldsymbol{\pi}^0$. □

Proof of Proposition 1. We prove statements (i), (ii) and (iii) in order. Part (i). To show that the minimizer of PA-C also minimizes the proxy loss function $Z_n(\boldsymbol{\pi})$, it suffices to show that solving PA-C and minimizing $Z_n(\boldsymbol{\pi})$ are equivalent problems in the following sense: (a) For any $(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ that is feasible for PA-C, $\boldsymbol{\pi} \in \Pi$; (b) For any $\boldsymbol{\pi} \in \Pi$, there exists $(\mathbf{x}, \mathbf{z})$ such that $(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ is feasible to PA-C; and (c) for any $(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ feasible for PA-C, $Z_n(\boldsymbol{\pi}) = Z_n^C(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$, where $Z_n^C(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ is the objective of PA-C. Statements (a) and (b) follow immediately by the construction of PA-C and the proxy loss problem (2.8); it remains to prove (c). Let $\boldsymbol{\pi}$ be fixed in PA-C, and let $(\mathbf{x}, \mathbf{z})$ be a solution to the resulting subproblem. Next, note

$$\begin{aligned}\sum_{i\in I}(\pi_{aj} - y_j^i)x_a^i &= \sum_{i\in\{i|x^i=a\}}(\pi_{aj} - y_j^i),\\ &= (\pi_{aj} - \omega_{aj})|\{i|x^i = a\}|,\\ &= (\pi_{aj} - \omega_{aj})\eta_{aj}, \end{aligned} \tag{A.21}$$

where the first equality follows because $x_a^i = 1$ if and only if $i \in \{i|x^i = a\}$, the second equality follows from the definition of $\boldsymbol{\omega}$, and the third equality follows by definition of $\boldsymbol{\eta}$. Then we have

$$\begin{aligned} Z_n^C(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z}) &= \sum_{a\in A}\sum_{j\in J}\Big|\frac{1}{n}\sum_{i\in I}(\pi_{aj} - y_j^i)x_a^i\Big|,\\ &= \frac{1}{n}\sum_{a\in A}\sum_{j\in J}\Big|(\pi_{aj} - \omega_{aj})\eta_{aj}\Big|,\\ &= Z_n(\boldsymbol{\pi}), \end{aligned}$$

where the first equality follows from (2.10a)−(2.10c), the second equality follows from (A.21), the third equality follows by definition of the element-wise norm $\|\cdot\|_1$, and the final equality follows by definition of $Z_n(\boldsymbol{\pi})$. Therefore, $Z_n(\boldsymbol{\pi}) = Z_n^C(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ for any $(\boldsymbol{\pi}, \mathbf{x}, \mathbf{z})$ feasible for PA-C. It follows that $\boldsymbol{\pi}_n^* \in \text{argmin}_{\boldsymbol{\pi}\in\Pi} Z_n(\boldsymbol{\pi})$. Part (ii). The proof of Theorem 1 establishes that $L_n(\hat{\boldsymbol{\pi}}_n) \longrightarrow 0$. Note $L_n(\hat{\boldsymbol{\pi}}_n) \leq L_n(\bar{\boldsymbol{\pi}}_n)$, because $\hat{\boldsymbol{\pi}}_n$ is a minimizer of $L_n(\boldsymbol{\pi})$ by definition. Therefore, by definition of $L_n(\boldsymbol{\pi})$, it suffices to show that for all (a, j) and $\varepsilon > 0$,

$$\Pr\left(\frac{1}{|I_a(\bar{\boldsymbol{\pi}}_n)|}\sum_{i\in I_a(\bar{\boldsymbol{\pi}}_n)} y_j^i - (\bar{\pi}_{aj})_n > \frac{\varepsilon}{md}\right) \longrightarrow 0, \tag{A.22}$$

where $I_a(\boldsymbol{\pi}) = \{i \in I | \mathbf{r}^i \in R_a(\boldsymbol{\pi})\}$. In the remainder of the proof, we suppress dependence of $\bar{\boldsymbol{\pi}}_n$ on n for conciseness. Next,

$$\begin{aligned}
\Pr\left(\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|}\sum_{i\in I_a(\bar{\boldsymbol{\pi}})} y_j^i - \bar{\pi}_{aj} > \frac{\varepsilon}{md}\right) &= \Pr\left(\sum_{i\in I_a(\bar{\boldsymbol{\pi}})} y_j^i - \bar{\pi}_{aj}|I_a(\bar{\boldsymbol{\pi}})| > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right), \\
&\leq \Pr\left(\sum_{b\in A}\sum_{j\in J}\Big|\sum_{i\in I_a(\bar{\boldsymbol{\pi}})} y_j^i - \bar{\boldsymbol{\pi}}_{aj}|I_a(\bar{\boldsymbol{\pi}})|\Big| > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right), \\
&= \Pr\left(nZ_n(\bar{\boldsymbol{\pi}}) > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right), \\
&\leq \Pr\left(nZ_n(\boldsymbol{\pi}^0) > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right),
\end{aligned}$$

where the first line follows by multiplying both sides of the inequality by $|I_a(\bar{\boldsymbol{\pi}})|$, the second line follows by non-negativity of the absolute value and summing over $b \in A$ and $j \in J$, the third line follows by definition of $Z_n(\boldsymbol{\pi})$, and the fourth line follows because $\bar{\boldsymbol{\pi}} \in \text{argmin}_{\boldsymbol{\pi}\in\bar{\Pi}} Z_n(\boldsymbol{\pi})$, by

definition. We can now write

$$\begin{aligned}
\Pr\left(nZ_n(\boldsymbol{\pi}^0) > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right) &= \Pr\left(\sum_{b\in A}\sum_{j\in J}\Big|\sum_{i\in I_b(\boldsymbol{\pi}^0)} y_j^i - \pi_{bj}^0|I_b(\boldsymbol{\pi}^0)|\Big| > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{md}\right) \\
&\leq \sum_{b\in A}\sum_{j\in J}\Pr\left(\Big|\sum_{i\in I_b(\boldsymbol{\pi}^0)} y_j^i - \pi_{bj}^0|I_b(\boldsymbol{\pi}^0)|\Big| > \frac{\varepsilon|I_a(\bar{\boldsymbol{\pi}})|}{m^2d^2}\right) \\
&\leq \sum_{b\in A}\sum_{j\in J}\sum_{k=0}^{n}\Pr\left(\Big|\sum_{i\in I_b(\boldsymbol{\pi}^0)} y_j^i - \pi_{bj}^0|I_b(\boldsymbol{\pi}^0)|\Big| > \frac{\varepsilon k}{m^2d^2}\middle||I_a(\bar{\boldsymbol{\pi}})| = k\right) \\
&\quad \times \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) \\
&\leq 2\sum_{b\in A}\sum_{j\in J}\sum_{k=0}^{n}\exp\left(-\left(\frac{\varepsilon k}{m^2d^2}\right)^2\cdot\frac{1}{|I_b(\boldsymbol{\pi}^0)|}\right)\Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) \\
&\leq 2d\sum_{b\in A}\sum_{k=0}^{n}\exp\left(-\left(\frac{\varepsilon k}{m^2d^2}\right)^2\cdot\frac{1}{n}\right)\Pr(|I_a(\bar{\boldsymbol{\pi}})| = k),
\end{aligned}$$

where the first line follows by definition of $Z_n(\boldsymbol{\pi})$, the second line follows from the union bound, the third line follows by conditioning on $|I_a(\bar{\boldsymbol{\pi}})|$, the fourth line follows by Hoeffding's inequality, and the fifth line follows by summing over $j \in J$ and because $n \geq |I_b(\boldsymbol{\pi}^0)|$ for all $b \in A$. Thus, to prove that (A.22) holds, it remains to show that for each $a \in A$,

$$\sum_{k=0}^{n}\exp\left(-\left(\frac{\varepsilon k}{m^2d^2}\right)^2\cdot\frac{1}{n}\right)\Pr(I_a(\bar{\boldsymbol{\pi}}) = k) \longrightarrow 0.$$

Next, using the fact that $\Pr(|I_a(\bar{\boldsymbol{\pi}})| = k)$, $k = 1, \ldots, n$ are binomial probabilities with parameter $\Pr(\mathbf{r} \in R_a(\bar{\boldsymbol{\pi}}))$, and $\Pr(\mathbf{r} \in R_a(\boldsymbol{\pi})) > 0$ for all $a \in A$ and $\boldsymbol{\pi} \in \Pi$ (Assumption 1), it can be shown with some effort that for any $\delta \in (0, 1)$, there exists $n > 0$ and $K \in [0, n]$ such that $\Pr(|I_a(\boldsymbol{\pi})| \leq K) \leq \delta$ for any $\boldsymbol{\pi} \in \Pi$ and $\exp(-(\frac{\varepsilon k}{m^2d^2})^2 \cdot \frac{1}{n}) < \delta$ for all $k \geq K$. It follows that

$$\sum_{k=0}^{K}\Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) = \Pr(|I_a(\bar{\boldsymbol{\pi}})| \leq K) \leq \delta \tag{A.23}$$

and

$$\sum_{k=K}^{n} \delta \cdot \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) = \delta \sum_{k=K}^{n} \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) \leq \delta. \tag{A.24}$$

Then we can write

$$\sum_{k=0}^{n} \exp\left(-\left(\frac{\varepsilon k}{m^2 d^2}\right)^2 \cdot \frac{1}{n}\right) \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) \leq \sum_{k=0}^{K-1} \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k) + \sum_{k=K}^{n} \delta \cdot \Pr(|I_a(\bar{\boldsymbol{\pi}})| = k)$$

$$\leq 2\delta,$$

where the first term after the first inequality follows because $\exp(-(\frac{\varepsilon k}{m^2 d^2})^2 \cdot \frac{1}{n}) \leq 1$ for all $k \in [0, K-1]$ and $n \geq 0$, and the second term follows because $\exp(-(\frac{\varepsilon k}{m^2 d^2})^2 \cdot \frac{1}{n}) \leq \delta$ for all $k \geq K$. Letting $\delta \longrightarrow 0$ yields the result. Part (iii). Because $|L_n(\boldsymbol{\pi}_n^*) - L_n(\hat{\boldsymbol{\pi}}_n)| \longrightarrow 0$ by part (ii) above, and $|L_n(\hat{\boldsymbol{\pi}}_n) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$ by Theorem 1, $|L_n(\boldsymbol{\pi}_n^*) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$. The remainder of the proof follows by parallel argument to the proof of Theorem 1, with $\boldsymbol{\pi}_n^*$ in place of $\hat{\boldsymbol{\pi}}_n$. □

A.5.2 Proof of Theorem 2

We first present two helpful supporting results in Lemmas 5 and 6. Lemma 5 is a concentration inequality that bounds the distance between an empirical mass function obtained from sampling from a discrete distribution and the discrete distribution itself. In Lemma 6, we define a Bernoulli random variable $e_a(\boldsymbol{\pi}^0, \boldsymbol{\pi})$, whose value depends on the realization of $\mathbf{r}$, and is equal to 1 if the agent's optimal action is a under the true parameter $\boldsymbol{\pi}^0$ but not under an alternative model $\boldsymbol{\pi}$. Intuitively, the event $\{e_a(\boldsymbol{\pi}^0, \boldsymbol{\pi}) = 1\}$ represents a "mis-classification" of the agent action by the model $\boldsymbol{\pi}$. Lemma 6 develops a bound on the probability that $e_a(\boldsymbol{\pi}^0, \boldsymbol{\pi})$ is positive (i.e., equal to 1), which is the key result that we use to obtain the bound in Theorem 2.

Lemma 5. *Let ξ^1, ξ^2, ..., ξ^n be i.i.d. discrete random variables with support $J = \{1, 2, \ldots, d\}$ and mass function $\lambda_j = \Pr(\xi = j)$ for $j \in J$. Define $f_j = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{\xi^i = j\}$ to be the empirical probability. Then*

$$\Pr\left(\sup_{j\in J} |f_j - \lambda_j| > \varepsilon\right) \leq 2\exp(-n\varepsilon^2).$$

Proof. For convenience, let $\Lambda_j = \sum_{k=1}^{j} \lambda_k$ be the cumulative distribution and let $F_j = \sum_{k=1}^{j} f_k$ be the empirical cumulative distribution. With some effort, it can be shown that $\sup_{j\in J} |f_j - \lambda_j| \leq 2\sup_{j\in J} |F_j - \Lambda_j|$. It follows that for any $\varepsilon > 0$,

$$\Pr\left(\sup_{j\in J} |f_j - \lambda_j| > \varepsilon\right) \leq \Pr\left(2\sup_{j\in J} |F_j - \Lambda_j| > \varepsilon\right) = \Pr\left(\sup_{j\in J} |F_j - \Lambda_j| > \varepsilon/2\right) \leq 2\exp(-n\varepsilon^2),$$

where the final inequality is the Dvoretzky−Kiefer−Wolfowitz inequality (Massart, 1990). □

Lemma 6. *Let Assumption 3 hold. Let $e_a(\boldsymbol{\pi}^0, \boldsymbol{\pi})$ be a Bernoulli random variable equal to 1 if the events $\{\mathbf{r} \in R_a(\boldsymbol{\pi}^0)\}$ and $\{r \notin R_a(\boldsymbol{\pi})\}$ both occur. Then there exists $\bar{\boldsymbol{\pi}} \in \Pi$, and constants $\delta_1 \in (0, 1)$ and $\delta_2 \in (0, 1)$ such that*

$$\Pr(e_a(\boldsymbol{\pi}^0, \boldsymbol{\pi}) > 0) \leq 4m(1 - \delta_1(1 - \delta_2))^n \tag{A.25}$$

for all $a \in A$.

Proof. The proof proceeds in four steps. In the first step, we construct $\bar{\pi}$. In the second and third steps, we prove two useful inequalities. In the fourth step, we prove the inequality (A.25) in the lemma statement. Step 1. By Assumption 3, for each $a \in A$, there exists a cluster $s(a) \in S$ such that $B_{s(a)} \in R_a(\boldsymbol{\pi}^0)$. Pick $s(a)$ accordingly for each $a \in A$, and set $w_{as} = 1$ for $s = s(a)$ and $w_{as} = 0$ for $s \in S \setminus s(a)$. Then let $\bar{\boldsymbol{\pi}}_a = \sum_{s\in S} \mathbf{v}_s w_{as}$, for $a \in A$. Note that by construction, $\bar{\boldsymbol{\pi}}_a = \mathbf{v}_{s(a)}$ for $a \in A$. Step 2. First, we define the following useful quantity, which will be used

throughout the proof of Theorem 2:

$$u_a(\mathbf{r}, \boldsymbol{\pi}) = \inf_{b \in A \setminus a} \left\{ \sum_{j \in J} (\pi_{aj} - \pi_{bj}) r_j + c_b - c_a \right\}.$$

Intuitively, $u_a(\mathbf{r}, \boldsymbol{\pi})$ represents the difference in agent utility between action a and the highest-utility action other than a. Based on this definition, note that an action a is optimal if and only if $u_a(\mathbf{r}, \boldsymbol{\pi}) \geq 0$. In this step, we show that there exists $\delta_1 \in (0,1)$ and $\delta_2 \in (0,1)$ such that $\Pr(\mathbf{r} \in B_s) \geq \delta_1$ for all $s \in S$, and

$$\int_{R_a(\boldsymbol{\pi}^0)} \exp\left(-k \left(\frac{u_a(\mathbf{r}, \boldsymbol{\pi}^0)}{2\bar{r}d} \right)^2 \right) f(\mathbf{r}) dr \leq \delta_2 \tag{A.26}$$

for all $k \geq 0$ and $a \in A$. Recall that $\bar{r} = \sup_{\mathbf{r} \in R} \|\boldsymbol{r}\|_0 < \infty$; that is, $\bar{r}$ is an upperbound on the largest agent payment r_j. First, for the existence of $\delta_1 \in (0,1)$, it follows from the continuity of $f(\mathbf{r})$ on R (Assumption 1) and $B_s \subseteq R$ that $\Pr(\mathbf{r} \in B_s) > 0$ for all $s \in S$. Letting $\delta_1 = \inf_{s \in S} \Pr(\mathbf{r} \in B_s)$ implies $\Pr(\mathbf{r} \in B_s) \geq \delta_1$ for all $s \in S$, as desired. Next, for the existence of $\delta_2 \in (0,1)$ such that (A.26) holds, first consider the case where $k = 0$. Then

$$\int_{R_a(\boldsymbol{\pi}^0)} \exp\left(-0 \left(\frac{u_a(\mathbf{r}, \boldsymbol{\pi}^0)}{2\bar{r}d} \right)^2 \right) f(\mathbf{r}) dr = \int_{R_a(\boldsymbol{\pi}^0)} f(\mathbf{r}) d\mathbf{r} < 1,$$

for all $a \in A$, where the final inequality follows because $R_a(\boldsymbol{\pi}^0) \subset R$ for all $a \in A$ (Assumption 1). It follows there exists $\tilde{\delta} \in (0,1)$ such that (A.26) holds for $k = 0$ and all $a \in A$. Now consider the case where $k \geq 1$. Observe that for all r such that $u_a(\mathbf{r}, \boldsymbol{\pi}^0) > 0$, $\exp(-(u_a(\mathbf{r}, \boldsymbol{\pi}^0)/(2\bar{r}d))^2) < 1$. Therefore, for each $a \in A$ there exists $\tilde{\delta}_a \in (0,1)$ such that if $u_a(\mathbf{r}, \boldsymbol{\pi}^0) > 0$, then $0 < \exp(-(u_a(\mathbf{r}, \boldsymbol{\pi}^0)/(2\bar{r}d))^2) \leq \tilde{\delta}_a < 1$. Next, exponentiating and integrating both sides of the preced-

ing inequality yields

$$\int_{R_a(\boldsymbol{\pi}^0)} \exp\left(-k\left(\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right)^2\right) f(\mathbf{r})d\mathbf{r} \leq \int_{R_a(\boldsymbol{\pi}^0)} (\tilde{\delta}_a)^k f(\mathbf{r})d\mathbf{r} \leq (\tilde{\delta}_a)^k.$$

The result follows from setting $\delta_2 = \sup\{\tilde{\delta}, \sup_{a\in A}\tilde{\delta}_a\}$. Step 3. In this step, we show that for $\bar{\pi}$ as constructed in step 1 and any $a \in A$,

$$\Pr\left(\sup_{j\in J}|\bar{\pi}_{bj} - \pi^0_{bj}| > \frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right) \leq 2(1-\delta_1(1-\delta_2))^n$$

for all $b \in A$. For convenience, let $I_s = \{i \in I | \mathbf{r}^i \in B_s\}$. Then for each $b \in A$,

$$\begin{aligned}
&\Pr\left(\sup_{j\in J}|\bar{\pi}_{bj} - \pi^0_{bj}| > \frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right) = \Pr\left(\sup_{j\in J}\left|\frac{1}{|I_{s(a)}|}\sum_{i\in I_{s(a)}} y^i_j - \pi^0_{bj}\right| > \frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right)\\
&= \sum_{k=0}^{n}\Pr\left(\sup_{j\in J}\left|\frac{1}{|I_{s(a)}|}\sum_{i\in I_{s(a)}} y^i_j - \pi^0_{bj}\right| > \frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\middle| |I_{s(a)}| = k\right)\Pr(|I_{s(a)}| = k)\\
&= \sum_{k=0}^{n}\left[\int_R \Pr\left(\sup_{j\in J}\left|\frac{1}{k}\sum_{i\in I_{s(a)}} y^i_j - \pi^0_{bj}\right| > \frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\middle| |I_{s(a)}| = k\right) f(\mathbf{r})dr\right]\Pr(|I_{s(a)}| = k)\\
&\leq 2\sum_{k=0}^{n}\left[\int_{R_a(\boldsymbol{\pi}^0)} \exp\left(-k\left(\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right)^2\right) f(\mathbf{r})dr\right]\Pr(|I_{s(a)}| = k)\\
&\leq 2\sum_{k=0}^{n}\tilde{\delta}^k \Pr(|I_{s(a)}| = k),
\end{aligned}$$

The first line follows because $\bar{\pi}_{aj} = v_{s(a),j} = (1/|I_{s(a)}|)\sum_{i\in I_{s(a)}} y^i_j$ for all (a,j), by construction of v_s (Algorithm 1) and $\bar{\boldsymbol{\pi}}$ (Step 1). The second line follows from the total probability rule, because $|I_{s(a)}|$ is a binomial random variable. The third line follows from conditioning on r and integrating over R, and because y^i_j for $i \in |I_{s(a)}|$ are independent of $\mathbf{r}$. The fourth line follows from Lemma 5, and because $R_a(\boldsymbol{\pi}^0) \subseteq R$. The fifth line follows from Step 2 of the proof. Next, note that $|I_s|$ is the number of observations that are contained in the ball B_s. Therefore, $|I_s|$ is a binomial random

variable with parameter $\lambda_s = \Pr(\mathbf{r} \in B_s)$. We can now write

$$\begin{aligned} 2\sum_{k=0}^{n} \delta_2^k \cdot \Pr(|I_{s(a)}| = k) &= 2\sum_{k=0}^{n} (\delta_2 \cdot \lambda_{s(a)})^k \frac{n!}{k!(n-k)!}(1-\lambda_{s(a)})^{n-k} \\ &= 2(1-\lambda_{s(a)}(1-\delta_2))^n \\ &\leq 2(1-\delta_1(1-\delta_2))^n, \end{aligned}$$

where the first equality follows from writing the binomial probabilities $\Pr(|I_{s(a)}| = k)$ explicitly and grouping terms raised to the k^{th} power, the second equality follows immediately from the binomial theorem, and the third inequality follows from Step 2. The result follows. Step 4. We now prove inequality (A.25). Note that by definition, $e_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}) > 0$ implies $r \notin R_a(\bar{\boldsymbol{\pi}})$, which implies there exists $b \in A$ such that $\sum_{j \in J}(\bar{\pi}_{bj} - \bar{\pi}_{aj})r_j > c_b - c_a$. Therefore, by the union bound,

$$\Pr(e_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}) > 0) \leq \sum_{b \in A \setminus a} \Pr\left(\sum_{j \in J}(\bar{\pi}_{bj} - \bar{\pi}_{aj})r_j > c_b - c_a\right).$$

In the remainder of the proof, we bound $\Pr\left(\sum_{j \in J}(\bar{\pi}_{bj} - \bar{\pi}_{aj})r_j > c_b - c_a\right)$ for each $b \in A \setminus a$. For each $b \in A \setminus a$, we have

$$\begin{aligned} &\Pr\left(\sum_{j \in J}(\bar{\pi}_{bj} - \bar{\pi}_{aj})r_j > c_b - c_a\right), \\ &\leq \Pr\left(\sum_{j \in J}(\bar{\pi}_{bj} - \pi^0_{bj})r_j - \sum_{j \in J}(\bar{\pi}_{aj} - \bar{\pi}^0_{aj})r_j > u_a(\mathbf{r}, \boldsymbol{\pi}^0)\right), \\ &\leq \Pr\left(\left\{\sum_{j \in J}(\bar{\pi}_{bj} - \pi^0_{bj})r_j > \frac{u_a(\mathbf{r}, \boldsymbol{\pi}^0)}{2}\right\} \cup \left\{\sum_{j \in J}(\bar{\pi}^0_{aj} - \bar{\pi}_{aj})r_j > \frac{u_a(\mathbf{r}, \boldsymbol{\pi}^0)}{2}\right\}\right), \end{aligned}$$

$$
\begin{aligned}
&\leq \Pr\left(\left\{\sup_{j\in J}|\bar{\pi}_{bj}-\pi^0_{bj}|>\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right\}\cup\left\{\sup_{j\in J}|\pi^0_{aj}-\bar{\pi}_{aj}|>\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right\}\right),\\
&\leq \Pr\left(\sup_{j\in J}|\bar{\pi}_{bj}-\pi^0_{bj}|>\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right)+\Pr\left(\sup_{j\in J}|\pi^0_{aj}-\bar{\pi}_{aj}|>\frac{u_a(\mathbf{r},\boldsymbol{\pi}^0)}{2\bar{r}d}\right),\\
&\leq 4(1-\delta_1(1-\delta_2))^n.
\end{aligned}
$$

The first inequality follows because $u_a(\mathbf{r},\boldsymbol{\pi}^0)\leq\sum_{j\in J}(\pi^0_{aj}-\pi^0_{bj})r_j+c_b-c_a$ for all $b\in A$, by definition of $u_a(\mathbf{r},\boldsymbol{\pi}^0)$. The second inequality follows because $\sum_{j\in J}(\pi_{bj}-\pi^0_{bj})r_j-\sum_{j\in J}(\bar{\pi}_{aj}-\pi^0_{aj})r_j>u_a(\mathbf{r},\boldsymbol{\pi}^0)$ implies at least one of $\sum_{j\in J}(\bar{\pi}_{bj}-\pi^0_{bj})r_j>u_a(\mathbf{r},\boldsymbol{\pi}^0)/2$ or $\sum_{j\in J}(\pi^0_{aj}-\bar{\pi}_{aj})r_j>u_a(\mathbf{r},\boldsymbol{\pi}^0)/2$ holds. The third inequality follows because $|J|\sup_{j\in J}\{(\boldsymbol{\pi}^0_{aj}-\bar{\boldsymbol{\pi}}_{aj})r_j\}\geq\sum_{j\in J}(\pi^0_{aj}-\bar{\pi}_{aj})r_j$, $|J|=d$, and $\bar{r}\geq r_j$ for all $j\in J$. The fourth inequality follows from the union bound. The fifth inequality follows from Step 3. The result follows by summing the final inequality over $b\in A$. □

Proof of Theorem 2. We prove the result by constructing a sequence of feasible solutions $\bar{\boldsymbol{\pi}}_n$ to PA-D, and bounding the objective function under $\bar{\boldsymbol{\pi}}_n$. For conciseness we suppress dependence of $\bar{\boldsymbol{\pi}}_n$ and $\tilde{\boldsymbol{\pi}}_n$ on n. By Assumption 3, for each $a\in A$, there exists a cluster $s(a)\in S$ such that $B_{s(a)}\in R_a(\boldsymbol{\pi}^0)$. Pick $s(a)$ accordingly for each $a\in A$. Set $\bar{w}_{as}=1$ for $s=s(a)$ and $\bar{w}_{as}=0$ for all $s\in S\setminus s(a)$. Fix $\mathbf{w}=\bar{\mathbf{w}}$ in PA-D and let $(\bar{\boldsymbol{z}},\bar{\boldsymbol{x}},\bar{\boldsymbol{\phi}})$ be the solution of the resulting subproblem. Then $(\bar{\mathbf{w}},\bar{\mathbf{z}},\bar{\mathbf{x}},\bar{\boldsymbol{\phi}})$ is a feasible solution to PA-D, and the associated estimate is given by

$$
\bar{\boldsymbol{\pi}}_a=\sum_{s\in S}\mathbf{v}_s\bar{w}_{as}=\mathbf{v}_{s(a)},\quad a\in A.
$$

Because $\bar{\boldsymbol{\pi}}$ is attained at a feasible solution to PA-D, $Z_n(\tilde{\boldsymbol{\pi}}) \leq Z_n(\bar{\boldsymbol{\pi}})$. It follows that

$$\Pr(|Z_n(\hat{\boldsymbol{\pi}}) - Z_n(\tilde{\boldsymbol{\pi}})| > \varepsilon) \leq \Pr(Z_n(\tilde{\boldsymbol{\pi}}) > \varepsilon) \leq \Pr(Z_n(\bar{\boldsymbol{\pi}}) > \varepsilon),$$

for any $\varepsilon > 0$. Therefore, it suffices to show that $\Pr(Z_n(\bar{\boldsymbol{\pi}}) > \varepsilon) = O(n^2\kappa^n)$ for some $\kappa \in (0,1)$. Because $Z_n(\boldsymbol{\pi}) = \sum_{a\in A}\sum_{j\in J}\frac{|I_a(\boldsymbol{\pi})|}{n}|\pi_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})|$ by definition, we shall bound $Z_n(\bar{\boldsymbol{\pi}})$ by bounding $\frac{|I_a(\boldsymbol{\pi})|}{n}|\bar{\pi}_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})|$ for each (a,j). Next, pick any (a,j). Then

$$\begin{aligned}
\frac{|I_a(\bar{\boldsymbol{\pi}})|}{n}\,|\bar{\pi}_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})| &= \frac{|I_a(\bar{\boldsymbol{\pi}})|}{n}\left|\bar{\pi}_{aj} - \pi^0_{aj} + \pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0) + \omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})\right|, \\
&\leq \frac{|I_a(\bar{\boldsymbol{\pi}})|}{n}\left(\left|\bar{\pi}_{aj} - \pi^0_{aj}\right| + \left|\pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0)\right| + \left|\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})\right|\right), \\
&\leq |\bar{\pi}_{aj} - \pi^0_{aj}| + |\pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0)| + |\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})|, \qquad \text{(A.27)}
\end{aligned}$$

where the second line follows by the triangle inequality, and the third line follows because $|I_a(\boldsymbol{\pi})| \leq n$ for any $\boldsymbol{\pi} \in \Pi$. The remainder of the proof proceeds in three steps. In each step, we bound one of the terms in the right-hand side of (A.27), from left to right. Step 1. For the first term, $|\bar{\pi}_{aj} - \pi^0_{aj}|$, note

$$\begin{aligned}
\Pr(|\bar{\pi}_{aj} - \pi^0_{aj}| > \varepsilon) &= \sum_{k=0}^{n} Pr(|\bar{\pi}_{aj} - \pi^0_{aj}| > \varepsilon \,||I_{s(a)}| = k)\Pr(|I_{s(a)}| = k) \\
&= \sum_{k=0}^{n} \Pr\left(\left|\frac{1}{|I_{s(a)}|}\sum_{i\in I_{s(a)}} y^i_j - \pi^0_{aj}\right| > \varepsilon \middle| |I_{s(a)}| = k\right)\Pr(|I_{s(a)}| = k) \\
&\leq \sum_{k=0}^{n} 2\exp(-\varepsilon^2 k)\Pr(|I_{s(a)}| = k).
\end{aligned}$$

The first line follows by conditioning on $|I_{s(a)}| = k$. The second line follows because $\bar{\pi}_{aj} = v_{s(a),j} = (1/|I_{s(a)}|)\sum_{i\in I_{s(a)}} y^i_j$ by the construction of $\bar{\pi}_{aj}$ above and the definition of $\mathbf{v}_{s(a)}$ (Algorithm 1). For the third line, note that $(1/|I_{s(a)}|)\sum_{i\in I_{s(a)}} y^i_j$ is the empirical mass function for the independent

variables ξ^i for all $i \in I_{s(a)}$, generated by $\boldsymbol{\pi}^0$. The inequality, therefore, follows by an application of Lemma 5. Next, because the $\mathbf{r}^i$ are i.i.d. (Assumption 1), $|I_{s(a)}|$ is a binomial random variable with parameter $\Pr(\mathbf{r} \in B_{s(a)})$. For conciseness, define $\lambda_{s(a)} = \Pr(\mathbf{r} \in B_{s(a)})$. We now have

$$\begin{aligned}\sum_{k=0}^{n} \left(2\exp(-\varepsilon^2 k)\right) \Pr(|I_{s(a)}| = k) &= \sum_{k=0}^{n} \left(2\exp(-\varepsilon^2 k)\right) \frac{n!}{k!(n-k)!} \lambda_{s(a)}^k (1-\lambda_{s(a)})^{n-k} \\ &= 2\sum_{k=0}^{n} \left(\lambda_{s(a)} \exp(-\varepsilon^2)\right)^k \frac{n!}{k!(n-k)!} (1-\lambda_{s(a)})^{n-k},\end{aligned}$$

where the first equality follows from writing the binomial probabilities $\Pr(|I_{s(a)}| = k)$ explicitly, and the second equality follows from grouping terms raised to the k^{th} power. Finally, we have

$$\begin{aligned}2\sum_{k=0}^{n} \left(\lambda_{s(a)} \exp(-\varepsilon^2)\right)^k \frac{n!}{k!(n-k)!} (1-\lambda_{s(a)})^{n-k} &= 2(1-\lambda_{s(a)}(1-\exp(-\varepsilon^2)))^n \\ &\leq 2(1-\delta_1(1-\exp(-\varepsilon^2)))^n,\end{aligned}$$

for some $\delta_1 \in (0,1)$, where the equality follows by applying the binomial theorem, and the inequality follows by letting $\delta_1 = \inf_{s \in S} \lambda_{s(a)}$. Therefore,

$$\Pr(|(\bar{\pi}_{aj} - \pi^0_{aj}| > \varepsilon) \leq 2(1-\delta_1(1-\exp(-\varepsilon^2)))^n. \tag{A.28}$$

Step 2. Next, we bound the second term $|\pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0)|$ in (A.27). Because

$$\omega_{aj}(\boldsymbol{\pi}^0) = (1/|I_a(\boldsymbol{\pi}^0)|) \sum_{i \in I_a(\boldsymbol{\pi}^0)} y^i_j,$$

it follows immediately from Lemma 5 that

$$\Pr(|\pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0)| > \varepsilon) \leq 2\exp(-n\varepsilon^2). \tag{A.29}$$

Step 3. Next, we bound the third term $|\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})|$ in (A.27). For convenience, let $\beta_{aj}(\boldsymbol{\pi})$ be a Bernoulli random variable equal to 1 if the events $\{i \in I_a(\boldsymbol{\pi})\}$ and $\{y_j = 1\}$ both occur, and let $\beta^i_{aj}(\boldsymbol{\pi})$ be the realized value of $\beta_{aj}(\boldsymbol{\pi})$ in the i^{th} observation. Using the definition of β^i_{aj}, it follows that

$$\begin{aligned}|\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})| &= \left| \frac{1}{|I_a(\boldsymbol{\pi}^0)|} \sum_{i \in I_a(\boldsymbol{\pi}^0)} y^i_j - \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I_a(\bar{\boldsymbol{\pi}})} y^i_j \right| \\ &= \left| \frac{1}{|I_a(\boldsymbol{\pi}^0)|} \sum_{i \in I} \beta^i_{aj}(\boldsymbol{\pi}^0) - \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} \beta^i_{aj}(\bar{\boldsymbol{\pi}}) \right|.\end{aligned}$$

Next, we have

$$\begin{aligned}&\left| \frac{1}{|I_a(\boldsymbol{\pi}^0)|} \sum_{i \in I} \beta^i_{aj}(\boldsymbol{\pi}^0) - \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} \beta^i_{aj}(\bar{\boldsymbol{\pi}}) \right| \qquad \text{(A.30)} \\ &\leq \left| \max \left\{ \frac{1}{|I_a(\boldsymbol{\pi}^0)|}, \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \right\} \sum_{i \in I} \beta^i_{aj}(\boldsymbol{\pi}^0) - \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} \beta^i_{aj}(\bar{\boldsymbol{\pi}}) \right| \\ &\leq \left| \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} (\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}})) + \left| \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|} \right| \sum_{i \in I} \beta^i_{aj}(\bar{\boldsymbol{\pi}}) \right| \\ &\leq \left| \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} (\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}})) \right| + \left| \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|} \right| \sum_{i \in I} \beta^i_{aj}(\bar{\boldsymbol{\pi}}), \qquad \text{(A.31)}\end{aligned}$$

where the first two lines are consequences of the $\max\{\cdot\}$ operator, and the third line follows from the triangle inequality. We shall bound each of the two terms on the right-hand side of (A.30) separately. For the first term on the right-hand side of (A.30), observe that by definition of $\beta^i_{aj}(\boldsymbol{\pi})$, $\sum_{j \in J} \beta^i_{aj}(\boldsymbol{\pi}) = 1$ if and only if $i \in I_a(\boldsymbol{\pi})$. Recall from Lemma 6 that $e_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}})$ is a Bernoulli variable equal to 1 if the events $\{\mathbf{r} \in R_a(\boldsymbol{\pi}^0)\}$ and $\{r \notin R_a(\bar{\boldsymbol{\pi}})\}$ both occur. Further, note $\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}}) > 0$ implies $\mathbf{r}^i \notin R_a(\bar{\boldsymbol{\pi}})$, which implies $e^i_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}) = 1$. Therefore,

$$\left| \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} (\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}})) \right| \leq \frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} \sum_{i \in I} e^i_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}).$$

Next, for any $\varepsilon \in (0,1)$,

$$\begin{aligned}\Pr\left(\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|}\sum_{i\in I} e_a^i(\boldsymbol{\pi}^0,\bar{\boldsymbol{\pi}}) > \varepsilon\right) &\leq \Pr\left(\sum_{i\in I} e_a^i(\boldsymbol{\pi}^0,\bar{\boldsymbol{\pi}}) > \varepsilon\right)\\ &\leq \sum_{i\in I}\Pr\left(e_a^i(\boldsymbol{\pi}^0,\bar{\boldsymbol{\pi}}) > \varepsilon\right)\\ &\leq n\Pr\left(e_a(\boldsymbol{\pi}^0,\bar{\boldsymbol{\pi}}) > \varepsilon\right),\end{aligned}$$

where the first inequality follows because $|I_a(\bar{\boldsymbol{\pi}})| \geq 1$ by Assumption 3 and the construction of $\bar{\boldsymbol{\pi}}$, the second inequality follows from applying the union bound, and the third inequality follows because the e_a^i are i.i.d. (Assumption 1). Next, by Lemma 6, $\Pr\left(e_a(\boldsymbol{\pi}^0,\bar{\boldsymbol{\pi}}) > \varepsilon\right) \leq 4m(1-\delta_1(1-\delta_2))^n$. Therefore, for any $\varepsilon \in (0,1)$,

$$\Pr\left(\left|\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|}\sum_{i\in I}(\beta_{aj}^i(\boldsymbol{\pi}^0) - \beta_{aj}^i(\bar{\boldsymbol{\pi}}))\right| > \varepsilon\right) \leq 4mn(1-\delta_1(1-\delta_2))^n. \tag{A.32}$$

Next, we bound the second term in (A.30). For any $\varepsilon \in (0,1)$, note

$$\Pr\left(\left|\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|}\right|\sum_{i\in I}\beta_{aj}^i(\bar{\boldsymbol{\pi}}) > \varepsilon\right) \leq n\Pr\left(\left|\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|}\right| > \varepsilon\right),$$

which follows from the union bound and because $\beta_{aj}^i(\bar{\boldsymbol{\pi}}) \leq 1$ for all $i \in I$. Next, note that

$$\left|\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|}\right| \leq ||I_a(\boldsymbol{\pi}^0)| - |I_a(\bar{\boldsymbol{\pi}})||,$$

which follows because $|I_a(\boldsymbol{\pi}^0)| \geq 1$ and $|I_a(\bar{\boldsymbol{\pi}})| \geq 1$ by Assumption 3 and by construction of $\bar{\boldsymbol{\pi}}$.

Therefore,

$$
\begin{aligned}
n \cdot \Pr\left(\left|\frac{1}{|I_a(\bar{\boldsymbol{\pi}})|} - \frac{1}{|I_a(\boldsymbol{\pi}^0)|}\right| > \varepsilon\right) &\leq n \cdot \Pr\left(||I_a(\boldsymbol{\pi}^0)| - |I_a(\bar{\boldsymbol{\pi}})|| > \varepsilon\right) \\
&= n \cdot \Pr\left(\left|\sum_{i\in I}\sum_{j\in J}\left(\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}})\right)\right| > \varepsilon\right) \\
&\leq n \cdot \Pr\left(\sum_{i\in I} e^i_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}) > \varepsilon\right) \\
&\leq n^2 \cdot \Pr\left(e_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}}) > \varepsilon\right) \\
&\leq 4mn^2(1 - \delta_1(1-\delta_2))^n, \qquad \text{(A.33)}
\end{aligned}
$$

where the second line follows because $|I_a(\boldsymbol{\pi})| = \sum_{i\in I}\sum_{j\in J}\beta^i_{aj}(\boldsymbol{\pi})$ by definition of $\beta^i_{aj}(\boldsymbol{\pi})$, the third line follows because $\sum_{j\in J}(\beta^i_{aj}(\boldsymbol{\pi}^0) - \beta^i_{aj}(\bar{\boldsymbol{\pi}})) \leq e^i_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}})$ for all $i \in I$, the fourth line follows by applying the union bound and noting that the $e^i_a(\boldsymbol{\pi}^0, \bar{\boldsymbol{\pi}})$ are i.i.d. (Assumption 1), and the fifth line follows from Lemma 6. Therefore, combining (A.32) and (A.33) produces the following bound on the third term in (A.27) for any $\varepsilon \in (0, 1)$:

$$
\Pr\left(|\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})| > \varepsilon\right) \leq 4mn(n+1)(1 - \delta_1(1-\delta_2))^n. \qquad \text{(A.34)}
$$

Combining (A.27), (A.28), (A.29) and (A.34) and applying the union bound yields

$$
\begin{aligned}
&\Pr\left(\frac{|I_a(\bar{\boldsymbol{\pi}})|}{n}|\bar{\pi}_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})| > \varepsilon\right) \\
&\leq \Pr\left(|\bar{\pi}_{aj} - \pi^0_{aj}| > \varepsilon/3\right) + \Pr\left(|\pi^0_{aj} - \omega_{aj}(\boldsymbol{\pi}^0)| > \varepsilon/3\right) + \Pr\left(|\omega_{aj}(\boldsymbol{\pi}^0) - \omega_{aj}(\bar{\boldsymbol{\pi}})| > \varepsilon/3\right) \\
&\leq 2\left(\exp(-n(\varepsilon/3)^2) + (1 - \delta_1(1 - \exp(-(\varepsilon/3)^2)))^n + 2mn(n+1)(1 - \delta_1(1-\delta_2))^n\right). \qquad \text{(A.35)}
\end{aligned}
$$

for any $\varepsilon \in (0, 1)$ and each $a \in A$ and $j \in J$. It follows that for any $\varepsilon \in (0, 1)$,

$$\Pr(Z_n(\bar{\boldsymbol{\pi}}) > \varepsilon) = \Pr\left(\sum_{a\in A}\sum_{j\in J}\frac{|I_a(\bar{\boldsymbol{\pi}})|}{n}|\bar{\pi}_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})| > \varepsilon\right) \tag{A.36}$$

$$\leq \sum_{a\in A}\sum_{j\in J}\Pr\left(|\bar{\pi}_{aj} - \omega_{aj}(\bar{\boldsymbol{\pi}})| > \frac{\varepsilon}{md}\right)$$

$$\leq 2md(\exp(-n(\varepsilon/(3md))^2) + \left(1 - \delta_1(1 - \exp(-(\varepsilon/(3md))^2))\right)^n \tag{A.37}$$

$$+ 2mn(n+1)(1 - \delta_1(1-\delta_2))^n)). \tag{A.38}$$

where the first line follows by definition of $Z_n(\boldsymbol{\pi})$, the second line follows from applying the union bound over all $a \in A$ and $j \in J$ and noting $m = |A|$ and $d = |J|$, and the third line follows from (A.35). Note that the third term in (A.38) is dominant, which implies $\Pr(Z_n(\bar{\boldsymbol{\pi}}) > \varepsilon) \leq O(n^2\kappa^n)$, where $\kappa = 1 - \delta_1(1-\delta_2)$. Lastly, note $\kappa \in (0,1)$ because $\delta_1 \in (0,1)$ and $\delta_2 \in (0,1)$. □

A.5.3 Proofs of Proposition 2, Theorem 3, and Theorem 4

Proof of Proposition 2. The proof proceeds in two steps. First, we show $Z_n(\tilde{\bar{\boldsymbol{\pi}}}_n) \longrightarrow 0$. Second, we prove the main result. Step 1. Note that for any $\bar{\boldsymbol{\pi}}$ attained at a feasible solution to PA-D(S), $0 \leq Z_n(\tilde{\bar{\boldsymbol{\pi}}}_n) \leq Z_n(\bar{\boldsymbol{\pi}})$. Therefore, we prove the result by constructing a feasible solution $(\bar{\boldsymbol{\pi}}, \bar{x}, \bar{w}, \bar{z}, \bar{\phi})$ for each $n \geq 0$ and showing $Z_n(\bar{\boldsymbol{\pi}}) \longrightarrow 0$. Because $Z_n(\boldsymbol{\pi}^0) \longrightarrow 0$ by Lemma 4, it suffices to show $\bar{\boldsymbol{\pi}} \longrightarrow \boldsymbol{\pi}_0$. By Assumption 3, for each $a \in A$ there exists $s \in S$ such that $B_s \subseteq R_a(\boldsymbol{\pi}^0)$. For each $n \geq 0$, let $(\bar{\boldsymbol{\pi}}, \bar{x}, \bar{w}, \bar{z}, \bar{\phi})$ be constructed as follows: for each $a \in A$, set $\bar{w}_{as} = 1$ for $s = s(a)$ and $\bar{w}_{as} = 0$ for all $s \in S \setminus s(a)$. Fix $w = \bar{w}$ in PA-D and let $(\bar{\boldsymbol{\pi}}, \bar{x}, \bar{z}, \bar{\phi})$ be the

solution of the resulting subproblem. Then

$$\begin{aligned}\Pr\left(\sup_{j\in J}|\bar{\pi}_{aj}-\pi^0_{aj}|>\varepsilon\right) &\leq \Pr\left(\sup_{j\in J}|(v_{s(a)j}-\pi^0_{aj}|>\varepsilon\right)\\ &\leq \Pr\left(\sup_{j\in J}\Big|\frac{1}{|I_{s(a)}|}\sum_{i\in I_{s(a)}} y^i_j-\pi^0_{aj}\Big|>\varepsilon\right)\\ &\leq 2\exp(-\varepsilon^2|I_{s(a)}|),\end{aligned}$$

where the first line follows from constraint (2.13i), the second line follows by definition of $\mathbf{v}_{s(a)}$, and the third line follows from Lemma 5. Because $f(\mathbf{r})$ is continuous on R (Assumption 1), $|I_{s(a)}| \longrightarrow \infty$ as $n \longrightarrow \infty$ for all $a \in A$. Therefore, $\bar{\boldsymbol{\pi}}_n \longrightarrow \boldsymbol{\pi}_0$, as desired. Step 2. Note $0 \leq Z_n(\boldsymbol{\pi}^*_n) \leq Z_n(\tilde{\boldsymbol{\pi}}_n)$ by defintion of $\boldsymbol{\pi}^*_n$. Because $Z_n(\tilde{\boldsymbol{\pi}}_n) \longrightarrow 0$ by Step 1, it follows that $Z_n(\boldsymbol{\pi}^*_n) \longrightarrow 0$. Therefore, $|Z_n(\boldsymbol{\pi}^*_n) - Z_n(\tilde{\boldsymbol{\pi}}_n)| \longrightarrow 0$, as desired. □

Proof of Theorem 3. Let $\tilde{w}$ be obtained at an optimal solution to PA-D, and define $\tilde{V} = \{\mathbf{v}_s \in V | \sum_{a\in A} \tilde{w}_{as} = 1\}$. Let $\tilde{S}$ index the candidate distributions in $\tilde{V}$. Let V^+_T denote the candidate distributions at termination of Algorithm 2, with index set S^+_T. For conciseness, we suppress dependence of $\tilde{V}, V^+_T$, $\tilde{S}$ and S^+_T on n. The proof proceeds in two steps. First, we show $|Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}^+_n)| \longrightarrow 0$. Second, we prove the main result. Step 1. Observe that if $\tilde{V} \subseteq V^+_T$, then $Z_n(\boldsymbol{\pi}^+_n) \leq Z_n(\tilde{\boldsymbol{\pi}}_n)$. By optimality of $\tilde{\boldsymbol{\pi}}_n$ with respect to PA-D(S), $Z_n(\tilde{\boldsymbol{\pi}}_n) \leq Z_n(\boldsymbol{\pi}^+_n)$. Thus $|Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}^+_n)| \longrightarrow 0$ if $\tilde{V} \subseteq V^+_T$. It therefore suffices to show that for any $\mathbf{v}_s \in \tilde{V}$, there exists $\mathbf{v}_{s'} \in V^+_T$ such that $|\mathbf{v}_s - \mathbf{v}_{s'}| \longrightarrow 0$. Suppose otherwise. Then for all $n \geq 0$, there exists $\mathbf{v}_s \in \tilde{V}$ such that $\mathbf{v}_s \neq \mathbf{v}_{s'}$ for all $\mathbf{v}_{s'} \in V^+_T$. By the weak law of large numbers, for each $s \in S$ there exists $\boldsymbol{\nu}_s \in \mathbb{R}^d$ such that $\mathbf{v}_s \longrightarrow \boldsymbol{\nu}_s$ as $n_s \longrightarrow \infty$. It follows that for all $n \geq 0$, there exists $s \in \tilde{S}$ such that $\boldsymbol{\nu}_s \neq \boldsymbol{\nu}_{s'}$ for all $s' \in S^+_T$. Note $v_{sj} = \psi_{sj}/(\sum_{j\in J}\psi_{sj})$. Then for all $n \geq 0$, there

exists $s \in \tilde{S}$ such that $H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) < 0$ and $\boldsymbol{\nu}_s \neq \boldsymbol{\nu}_s{}'$ for all $s' \in S_T^+$. Next, it is straightforward to verify that because the density function $f(\mathbf{r})$ is continuous on R (Assumption 1), $n_s \longrightarrow \infty$ and $n_{s'} \longrightarrow \infty$ as $n \longrightarrow \infty$. Hence, $\Pr(H_\alpha(\boldsymbol{\psi}_s, \boldsymbol{\psi}_{s'}) \geq 0 | \boldsymbol{\nu}_s \neq \boldsymbol{\nu}_s{}') \longrightarrow 1$, a contradiction. Therefore, $|Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^+)| \longrightarrow 0$. Step 2. By Step 1 of the proof of Proposition 2, we have $Z_n(\tilde{\boldsymbol{\pi}}_n) \longrightarrow 0$. Because $|Z_n(\tilde{\boldsymbol{\pi}}_n) - Z_n(\boldsymbol{\pi}_n^+)| \longrightarrow 0$ from the first step of this proof, it follows that $Z_n(\boldsymbol{\pi}_n^+) \longrightarrow 0$. Following a parallel argument to the proof of Proposition 1, it can be shown that $Z_n(\boldsymbol{\pi}_n^+) \longrightarrow 0$ implies $|L_n(\boldsymbol{\pi}_n^+) - L_n(\hat{\boldsymbol{\pi}}_n)| \longrightarrow 0$. Because $|L_n(\hat{\boldsymbol{\pi}}_n) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$ from the first step of the proof of Theorem 1, $|L_n(\boldsymbol{\pi}_n^+) - L_n(\hat{\boldsymbol{\pi}}_n)| \longrightarrow 0$ implies $|L_n(\boldsymbol{\pi}_n^+) - L_n(\boldsymbol{\pi}^0)| \longrightarrow 0$. The remainder of the proof follows by a parallel argument to the second step of the proof of Theorem 1, with $\boldsymbol{\pi}_n^+$ in place of $\hat{\boldsymbol{\pi}}_n$. □

Proof of Theorem 4. The proof proceeds in two steps. First, we show $\Pr(T > m) \leq \alpha m S$. Second, we prove the main result. Step 1. By assumption, for each $s \in S$, there exists $a \in A$ such that $B_s \subseteq R_a(\boldsymbol{\pi}^0)$. In words, this means that for each ball B_s, every contract $\mathbf{r} \in B_s$ induces the same action from the agent. Accordingly, each candidate distribution $s \in S$ is *mapped* to exactly one action $a \in A$. If there exists an $s \in S^+$ such that s is mapped to an action a, we say that the action a is *represented* in S^+. For each $s \in S^-$, let E_s be the event that there exists $s'(s) \in S^+$ such that $B_s \subseteq R_a(\boldsymbol{\pi}^0)$ and $B_{s'} \subseteq R_a(\boldsymbol{\pi}^0)$; that is, E_s is the event that the action a that $s \in S^-$ maps to is already represented in S^+. Next, pick an iteration t, and let s_t^* be the cluster selected in the t^{th} iteration of Algorithm 2. Then $E_{s_t^*}$ is the event that the candidate distribution selected in iteration t is mapped to an action that is already represented in S^+. We can now write

$$
\begin{aligned}
\Pr(E_{s_t^*}) &\leq \Pr\left(\cup_{s\in S^-}\left\{\cap_{s'\in S^+}\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\},E_s\right\}\right)\\
&\leq \sum_{s\in S^-}\Pr\left(\cap_{s'\in S^+}\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\},E_s\right)\\
&\leq \sum_{s\in S^-}\Pr\left(\cap_{s'\in S^+}\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\}\Big|E_s\right)\Pr(E_s)\\
&\leq \sum_{s\in S^-}\Pr\left(\cap_{s'\in S^+}\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\}\Big|E_s\right), \qquad \text{(A.39)}
\end{aligned}
$$

The first inequality follows by definition of s^τ (Algorithm 2) and E_s. The second inequality follows from the union bound, the third inequality follows from conditioning on E_s, and the fourth inequality follows because $\Pr(E_s)\leq 1$ for all $s\in S$. Next, for each $s\in S$, we have

$$
\begin{aligned}
\Pr\left(\cap_{s'\in S^+}\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\}\Big|E_s\right) &\leq \Pr\left(H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'(s)})\geq 0\Big|E_s\right) &\text{(A.40)}\\
&\leq \Pr\left(H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'(s)})\geq 0\Big|\nu_s=\nu_{s'(s)}\right) &\text{(A.41)}\\
&\leq \alpha, &\text{(A.42)}
\end{aligned}
$$

where the first inequality follows from dropping the events $\{H_\alpha(\boldsymbol{\psi}_s,\boldsymbol{\psi}_{s'})\geq 0\}$ from the intersection for all elements in $S^+\setminus s'(s)$, the second inequality follows because the event $\nu_s=\nu_{s'(s)}$ implies the event E_s, and the third inequality follows by definition of $H_\alpha(\psi_s,\psi_{s'})$. By combining (A.39) and (A.40) and noting $|S^-|\leq|S|$ for all $t=1,\ldots,T$, it follows that $\Pr(E_{s_t^*})\leq\alpha|S|$. Next, note that the event $\{T>m\}$ implies $\{\cup_{t=0}^{m-1}E_{s_t^*}\}$ by the pigeonhole principle. Therefore, by the union bound,

$$\begin{aligned}\Pr(T > m) &\leq \Pr\left(\cup_{k=0}^{m-1} E_{s_t^*}\right) \\ &\leq \sum_{k=0}^{m-1} \Pr(E_{s_t^*}) \\ &\leq \sum_{k=0}^{m-1} \alpha|S| \\ &\leq \alpha m|S|,\end{aligned}$$

as desired. Step 2. Note $\mathbb{E}[T] = \sum_{k=0}^{\infty} \Pr(T > k)$ by definition of the expectation. It follows that

$$\begin{aligned}\sum_{k=0}^{\infty} \Pr(T > k) &= \sum_{k=0}^{|S|-1} \Pr(T > k) \\ &= \sum_{k=0}^{m-1} \Pr(T > k) + \sum_{k=m}^{|S|-1} \Pr(T > k) \\ &\leq m + \sum_{k=m}^{|S|-1} \Pr(T > k),\end{aligned}$$

where the first equality follows because T is bounded above by $|S|$, the second equality follows from separating the summation, and the inequality follows because $\sum_{k=0}^{m-1} \Pr(T > k) \leq m$. Next, note that the event $T > k$ implies $T > m$ for $k = m, \ldots, |S| - 1$, and thus $\Pr(T > k) \leq \Pr(T > m)$ for $k = m, \ldots, |S| - 1$. Therefore,

$$m + \sum_{k=m}^{|S|-1} \Pr(T > k) \leq m + \sum_{k=m}^{|S|-1} \Pr(T > m) = m + \Pr(T > m)(|S| - m) \leq m[1 + \alpha \cdot |S| \cdot (|S| - m)],$$

where the final inequality follows because $\Pr(T > m) \leq \alpha m|S|$ as established in Step 1. $\square$

A.5.4 Proofs for Sections A.3 and A.4

Proof of Corollary 1. The proof proceeds similarly to the proof of Theorem 1: First, we show $|\sum_{k\in K}\left(L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)\right)| \longrightarrow 0$. Second, we show $\text{plim}_{n\to\infty}\hat{\boldsymbol{\pi}}_n = \boldsymbol{\pi}^0$ if and only if Assumption 8 holds. Step 1. By Lemma 3, $L_n^k(\boldsymbol{\pi})$ is lower semicontinuous for all $k \in K$. Because the sum of lower semicontinuous functions is lower semicontinuous (Dietze and Schäuble, 1985), it follows that $\sum_{k\in K} L_n^k(\boldsymbol{\pi})$ is lower semicontinuous. Next, we show $|\sum_{k\in K}\left(L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)\right)| \longrightarrow 0$. By the triangle inequality,

$$\left|\sum_{k\in K}\left(L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)\right)\right| \leq \sum_{k\in K}|L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)|.$$

By the proof of Theorem 1 and Assumption 7, we have $|L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)| \longrightarrow 0$. Because $L_n^k(\boldsymbol{\pi}) \geq 0$ for any $\boldsymbol{\pi}$, it follows that $|\sum_{k\in K}\left(L_n^k(\hat{\boldsymbol{\pi}}_n) - L_n^k(\boldsymbol{\pi}^0)\right)| \longrightarrow 0$. Step 2. For each $k \in K$, define $L^k(\boldsymbol{\pi}) = \sum_{a\in A}\sum_{j\in J}|\pi_{aj}^k - \lim_{n\to\infty}\omega_{aj}^k(\boldsymbol{\pi})|$. Because $\sum_{k\in K} L_n^k(\boldsymbol{\pi})$ is lower semicontinuous, $|\sum_{k\in K} L_n^k(\hat{\boldsymbol{\pi}}_n) - \sum_{k\in K} L_n^k(\boldsymbol{\pi}^0)| \longrightarrow 0$, and Π is compact, by Theorem 5.14 of Van der Vaart (2000),

$$\text{plim}_{n\to\infty}\hat{\boldsymbol{\pi}}_n \in \underset{\pi\in\Pi}{\text{argmin}} \sum_{k\in K} L^k(\boldsymbol{\pi}).$$

Next, by parallel argument to Lemma 4, it can be shown that $\boldsymbol{\pi}^0$ is the unique minimizer of $\sum_{k\in K} L^k(\boldsymbol{\pi})$ if and only if Assumption 8 holds. Suppose Assumption 8 holds. It follows that $L^k(\boldsymbol{\pi}^0) = 0$ and $L^k(\boldsymbol{\pi}) > 0$ for $\boldsymbol{\pi} \neq \boldsymbol{\pi}^0$. Therefore, $\underset{\boldsymbol{\pi}\in\Pi}{\text{argmin}} \sum_{k\in K} L_n^k(\boldsymbol{\pi}) = \boldsymbol{\pi}^0$, which implies $\text{plim}_{n\to\infty}\hat{\boldsymbol{\pi}}_n = \boldsymbol{\pi}^0$. Conversely, if Assumption 8 does not hold, then there exists $\tilde{\boldsymbol{\pi}} \neq \boldsymbol{\pi}^0$ such that $\sum_{k\in K} L(\tilde{\boldsymbol{\pi}}) = \sum_{k\in K} L(\boldsymbol{\pi}^0) = 0$, which implies $\text{plim}_{n\to\infty}\hat{\boldsymbol{\pi}}_n \neq \boldsymbol{\pi}^0$. □

Proof of Proposition 5. For convenience, let $Z^t(\mathbf{w}^t, \boldsymbol{\varepsilon}^t)$ be the objective function value of PA-T under $(\mathbf{w}^t, \boldsymbol{\varepsilon}^t)$, and let $(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t)$ be the optimal solution to PA-T in iteration t of Algorithm 7. Let

"$\longrightarrow$" denote convergence in probability as $T \longrightarrow \infty$. The proof proceeds in four steps. We prove useful supporting results in the first three steps: In Step 1 we show $\mathbf{v}_s^T \longrightarrow \boldsymbol{\pi}^0_{a(\mathbf{r}^s)}$ for each $s \in S$; in Step 2 we show $Z^T(\bar{\mathbf{w}}^T, \bar{\boldsymbol{\varepsilon}}^T) \longrightarrow 0$; in Step 3 we show $\Pr(\bar{w}_{as}^T = 1) \longrightarrow 1$ for $a = a(\mathbf{r}^s)$ and $\Pr(\bar{w}_{as}^T = 0) \longrightarrow 1$ for $a \neq a(\mathbf{r}^s)$, where $a(\mathbf{r}^s)$ is the agent's true optimal action under $\mathbf{r}^s$ and $\boldsymbol{\pi}^0$. In Step 4, we combine these results to prove the statement of Proposition 5. Step 1. Because $\epsilon^t > 0$ for each $t \geq 1$, by Algorithm 7 we have $n_s^T \longrightarrow \infty$ as $T \longrightarrow \infty$. It follows by construction of $\mathbf{v}_s^t$ in Algorithm 7 and by Lemma 5 that $\mathbf{v}_s^T \longrightarrow \boldsymbol{\pi}^0_{a(\mathbf{r}^s)}$. Step 2. For each $t \geq 1$, construct a solution $\tilde{\mathbf{w}}$ as follows: for each $s \in S$, set $\tilde{w}_{as} = 1$ for $a = a(\mathbf{r}^s)$ and $\tilde{w}_{as} = 0$ for $a \neq a(\mathbf{r}^s)$. Note that $\tilde{\mathbf{w}}$ satisfies constraints (A.13c)-(A.13e) for all $t \geq 1$. Next, fix $\tilde{\mathbf{w}}$ as a parameter in PA-T. Then for each t, PA-T simplifies to the following subproblem:

$$\underset{\boldsymbol{\varepsilon}}{\text{minimize}} \quad \sum_{s \in S} |\varepsilon_s| \tag{A.43a}$$

$$\text{subject to} \quad \left(\sum_{j \in J} v_{sj}^t r_j^s - c_{a(\mathbf{r}^s)} \right) + \varepsilon_s \geq \left(\sum_{j \in J} v_{s'j}^t r_j^s - c_{a(\mathbf{r}^{s'})} \right), \quad s \in S, s' \in S, \tag{A.43b}$$

$$\varepsilon_s \geq 0, \quad s \in S. \tag{A.43c}$$

Let $\tilde{\boldsymbol{\varepsilon}}^t$ be the optimal solution to the above subproblem in round t. It follows from constraint (A.43b) and (A.43c) that for each $s \in S$,

$$\tilde{\varepsilon}_s^t = \max_{s' \in S} \left\{ \left(\sum_{j \in J} v_{s'j}^t r_j^s - c_{a(\mathbf{r}^{s'})} \right) - \left(\sum_{j \in J} v_{sj}^t r_j^s - c_{a(\mathbf{r}^s)} \right), 0 \right\}. \tag{A.44}$$

Note that $\sum_{j \in J} \pi^t_{a(\mathbf{r}^{s'}),j} r_j^s - c_{a(\mathbf{r}^{s'})} \leq \sum_{j \in J} \pi^t_{a(\mathbf{r}^s),j} r_j^s - c_{a(\mathbf{r}^s)}$ for all $k \in S$ and $t \geq 1$, by optimality of $a(\mathbf{r}^s)$ with respect to contract $\mathbf{r}^s$. Because $\mathbf{v}_s^T \longrightarrow \boldsymbol{\pi}^0_{a(\mathbf{r}^s)}$ for each $s \in S$ by Step 1, it follows from (A.44) that $\bar{\varepsilon}_s^T \longrightarrow 0$ for each $s \in S$. Therefore, $Z^T(\tilde{\mathbf{w}}, \tilde{\boldsymbol{\varepsilon}}^T) \longrightarrow 0$. By optimality of $(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t)$ with respect to PA-T, we have $Z^t(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t) \leq Z^t(\tilde{\mathbf{w}}, \tilde{\boldsymbol{\varepsilon}}^t)$ for all $t \geq 1$. Therefore, $Z^T(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t) \longrightarrow 0$.

Step 3. For each $s \in S$ and $t \geq 1$, let $\bar{a}^t(\mathbf{r}^s)$ be the action for which $\bar{\mathbf{w}}^t_{as} = 1$. Suppose by way of contradiction that $\Pr(\bar{a}^T(\mathbf{r}^s) = a(\mathbf{r}^s)) \longrightarrow 1$ does not hold for each $s \in S$. Then there exists $\tilde{s} \in S$ and a subsequence t_b, $b = 1, 2, \ldots$ such that $\bar{a}^{t_b}(\mathbf{r}^{\tilde{s}}) \neq a(\mathbf{r}^{\tilde{s}})$ for all $b \geq 1$. Next, by constraint (A.13b), we have

$$\begin{aligned}\varepsilon^{t_b}_{\tilde{s}} &\geq \max_{s' \in S} \left\{ \sum_{j \in J} (v^t_{s'j} - v^t_{sj}) r^{\tilde{s}}_j + c_{\bar{a}(\mathbf{r}^{\tilde{s}})} - c_{\bar{a}(\mathbf{r}^{s'})}, 0 \right\} \\ &\geq \max_{s' \in S} \left\{ \sum_{j \in J} (v^t_{s'j} - v^t_{sj}) r^{\tilde{s}}_j + \underline{c} - \bar{c}, 0 \right\}\end{aligned}$$

for all $b \geq 1$, where the second inequality follows by definition of $\underline{c}$ and $\bar{c}$. Because $\mathbf{v}^{t_b}_{s'} \longrightarrow \pi^0_{a(\mathbf{r}^{s'})}$ as $b \longrightarrow \infty$ for all $s' \in S$ by Step 1, and by Assumption 9 there exists $s' \in S$ such that $\sum_{j \in J} (\pi^0_{a(\mathbf{r}^{s'}),j} - \pi^0_{a(\mathbf{r}^{\tilde{s}}),j}) r^{\tilde{s}}_j + \underline{c} - \bar{c} > 0$, it follows that $\lim_{b \to \infty} \varepsilon^{t_b}_{\tilde{s}} > 0$. Hence $\lim_{b \to \infty} Z^{t_b}(\bar{\mathbf{w}}^{t_b}, \bar{\boldsymbol{\varepsilon}}^{t_b}) > 0$, which implies $\lim_{T \to \infty} Z^T(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t) > 0$. However, by Step 2, $Z^T(\bar{\mathbf{w}}^t, \bar{\boldsymbol{\varepsilon}}^t) \longrightarrow 0$, which yields a contradiction. Therefore, $\Pr(\bar{a}^T(\mathbf{r}^s) = a(\mathbf{r}^s)) \longrightarrow 1$ for all $s \in S$. By definition of $\bar{a}^t(\mathbf{r}^s)$, this implies $\Pr(\bar{w}^T_{as} = 1) \longrightarrow 1$ for $a = a(\mathbf{r}^s)$ and $\Pr(\bar{w}^T_{as} = 0) \longrightarrow 1$ for $a \neq a(\mathbf{r}^s)$, as desired. Step 4. We now show $\hat{\boldsymbol{\pi}}^T \longrightarrow \boldsymbol{\pi}^0$. For convenience, let $S_a = \{s | a(\mathbf{r}^s) = a\}$. Then by construction of $\hat{\boldsymbol{\pi}}^t$ from Algorithm 7,

$$\hat{\pi}^t_{aj} = \frac{\sum_{s \in S} v^t_{sj} n^t_s \bar{w}^t_{as}}{\sum_{s \in S} n^t_s \bar{w}^t_{as}} = \frac{\sum_{s \in S_a} v^t_{sj} n^t_s \bar{w}^t_{as} + \sum_{s \in S \setminus S_a} v^t_{sj} n^t_s \bar{w}^t_{as}}{\sum_{s \in S_a} n^t_s \bar{w}^t_{as} + \sum_{s \in S \setminus S_a} n^t_s \bar{w}^t_{as}},$$

for all (a, j). By Step 3, $\Pr(\bar{w}^T_{as} = 1) \longrightarrow 1$ for $s \in S_a$ and $\Pr(\bar{w}^T_{as} = 0) \longrightarrow 1$ for $s \in S \setminus S_a$. Further, by Step 1, $\mathbf{v}^T_s \longrightarrow \boldsymbol{\pi}^0_{a(\mathbf{r}^s)}$. It follows that for each $j \in J$, $v^T_{sj} \bar{w}^T_{as} \longrightarrow \pi^0_{aj}$ for $s \in S_a$, and $v^T_{sj} \bar{w}^T_{as} \longrightarrow 0$ for $s \in S \setminus S_a$. Therefore,

$$\operatorname*{plim}_{T \to \infty} \hat{\pi}^T_{aj} = \operatorname*{plim}_{T \to \infty} \frac{\sum_{s \in S_a} v^T_{sj} n^T_s \bar{w}^T_{as} + \sum_{s \in S \setminus S_a} v^T_{sj} n^T_s \bar{w}^T_{as}}{\sum_{s \in S_a} n^T_s \bar{w}^T_{as} + \sum_{s \in S \setminus S_a} n^T_s \bar{w}^T_{as}} = \operatorname*{plim}_{T \to \infty} \frac{\pi^0_{aj} \sum_{s \in S_a} n^T_s}{\sum_{s \in S_a} n^T_s} = \pi^0_{aj}.$$

for all (a, j). Because $\hat{\pi}_{aj}^T \longrightarrow \pi_{aj}^0$ for all (a, j), it follows that $\hat{\boldsymbol{\pi}}^T \longrightarrow \boldsymbol{\pi}^0$. □

Proof of Corollary 2. It suffices to show $|U(\hat{\mathbf{r}}^T) - U(\mathbf{r}^*)| \longrightarrow 0$. The proof proceeds in two steps. First, we show $\hat{U}(\mathbf{r}) \longrightarrow U(\mathbf{r})$ for any $\mathbf{r} \in R$. Second, we prove the main result. Step 1. Note

$$\begin{aligned} |U(\hat{\mathbf{r}}) - U(\mathbf{r}^*)| &= |U(\hat{\mathbf{r}}) - \hat{U}(\hat{\mathbf{r}}) + \hat{U}(\hat{\mathbf{r}}) - \hat{U}(\mathbf{r}^*) + \hat{U}(\mathbf{r}^*) - U(\mathbf{r}^*)| \\ &\leq |U(\hat{\mathbf{r}}) - \hat{U}(\hat{\mathbf{r}})| + |\hat{U}(\hat{\mathbf{r}}) - \hat{U}(\mathbf{r}^*)| + |\hat{U}(\mathbf{r}^*) - U(\mathbf{r}^*)| \end{aligned}$$

where the equality follows from adding and subtracting $\hat{U}(\hat{\mathbf{r}})$ and $\hat{U}(\mathbf{r}^*)$, and the inequality follows from the triangle inequality. Next, note that by the definitions of $\hat{U}(\mathbf{r})$ (given above) and $U(\mathbf{r})$ (given in (A.11)), $\hat{\boldsymbol{\pi}}^T \longrightarrow \boldsymbol{\pi}^0$ implies $\hat{U}(\mathbf{r}) \longrightarrow U(\mathbf{r})$ for any $\mathbf{r} \in R$. Step 2. By Step 1, $|U(\hat{\mathbf{r}}^T) - \hat{U}(\hat{\mathbf{r}}^T)| \longrightarrow 0$ and $|\hat{U}(\mathbf{r}^*) - U(\mathbf{r}^*)| \longrightarrow 0$. It remains to show that $|\hat{U}(\hat{\mathbf{r}}^T) - \hat{U}(\mathbf{r}^*)| \longrightarrow 0$. Suppose not. Then there exists a subsequence t_b, $b \geq 1$ such that $|\hat{U}(\hat{\mathbf{r}}^{t_b}) - \hat{U}(\mathbf{r}^*)| > 0$ for all $b \geq 1$. Further, because for all $t \geq 1$, $\hat{\mathbf{r}}^t \in \text{argmax}_{\mathbf{r} \in R} \hat{U}(\mathbf{r})$ by definition, we have $\hat{U}(\hat{\mathbf{r}}^{t_b}) > \hat{U}(\mathbf{r}^*)$ for all $b \geq 1$. Letting $b \longrightarrow \infty$ and noting $\hat{U}(\mathbf{r}) \longrightarrow U(\mathbf{r})$ for all $\mathbf{r} \in R$ yields $\text{plim}_{b \to \infty} \hat{U}(\hat{\mathbf{r}}^{t_b}) > U(\mathbf{r}^*)$. It is straightforward to obtain a contradiction to the preceding inequality using $\hat{U}(\mathbf{r}) \longrightarrow U(\mathbf{r})$ from Step 1, and $\mathbf{r}^* = \text{argmax}_{\mathbf{r} \in R} U(\mathbf{r})$ by definition of $\mathbf{r}^*$. □

Appendix B

Discovering Causal Models with Optimization

B.1 Finitely Many Paths

We provide here a more complete definition of paths in directed mixed graphs, which may contain paths with repeating nodes due to cycles. In the lemma that follows, we show that such paths can be eliminated from consideration without altering d-separation relations between nodes. This result serves as justification for the simplified path definition given in Definition 2.

Definition 8 (Path)**.** *Given a node set V, a set of edge-types $T = \{\rightarrow, \leftarrow, \leftrightarrow\}$ and an edge set E of triples (v_1, t, v_2) with $v_1, v_2 \in V$ and $t \in T$, we define a path p_{ij} from node i to node j with $i, j \in V, i \neq j$, as a sequence of edges $p_{ij} = (e_1, \ldots, e_\ell)$ such that*

(i) the edges in the path are in the edge set: $e_k \in E$ for all $1 \leq k \leq \ell$,

(ii) the path starts with node i: $e_1 = (i, t, v)$ for some $v \in V \setminus \{i\}$ and $t \in T$,

(iii) the path ends with node j: $e_\ell = (v, t, j)$ for some $v \in V \setminus \{j\}$ and $t \in T$,

(iv) consecutive edges on the path are connected: for all $e_k, e_{k+1} = (v_1, t, v_2)(u_1, t', u_2) \in p_{ij}$ with $1 \leq k < \ell$, we have $v_2 = u_1$.

A *directed path* from i to j is a path that only has edge-type $T = \{\rightarrow\}$, i.e. all edges point away from i and towards j along the path. So, a node j is a *descendant* of i if there is a directed path from i to j.

We say that an occurrence of a node $v \in V \setminus \{i, j\}$ on a path p_{ij} is when v is the endpoint of one edge and the starting point of the subsequent edge. In addition, i and j occur once at the beginning and end of the path, respectively. A node repeats on a path if it occurs more than once on the path. In that case the path is said to contain a cycle.

We now show that with respect to the d-separation and d-connection relations, we can ignore cycles on a path. That is, if a path with a cycle is d-connecting, then there is a path without the cycle that is also d-connecting.

Lemma 7. *Let path p_{ij} be a path between nodes $i, j \in V$ according to Definition 8. If path p_{ij} is unblocked with respect to conditioning set $C \subseteq V \setminus \{i, j\}$ in directed mixed graph $\mathcal{G} = (V, E)$ and has repeating nodes, then there exists a path p^*_{ij} without any repeating nodes in $\mathcal{G}$ that is also unblocked with respect to C.*

Proof of Lemma 7. Suppose there is a path p with repeating nodes between variables $i, j \in V$ in graph $\mathcal{G} = (V, E)$ that is unblocked with respect to conditioning set $C \subseteq V \setminus \{i, j\}$. Following from Definition 3, every collider k on the path p is in C or has a descendant in C, and no other nodes on the path are in C. Note that since path p is unblocked with respect to C, Definition 3 implies that the same node cannot be a collider and a noncollider at the same time on path p.

Now we show we can find a shorter path p^* in $\mathcal{G}$ that is also unblocked with respect to C. Note that path p includes the same node more than once by construction. Without loss of generality, let node l repeat on path p more than once. Let l_1 represent the *first* occurrence of node l on path p and let l_2 represent the *last* occurrence of node k on path p. Let p_{i-l_1} represent the subpath between i and l_1 on path p. Similarly, let p_{l_2-j} represent the subpath between l_2 and j on path p.

Note that both the last node on path p_{i-l_1} and the first node on path p_{l_2-j} are node l. Hence, we can obtain a path p^* where node l does not repeat by combining p_{i-l_1} and p_{l_2-j} together.

Next we show that path p^* is unblocked with respect to C. To do so, we need to show every collider k on the path p^* is in C or has a descendant in C, and no other nodes on the path p^* are in C by Definition 3. Notice that if node $k \neq l$ is a collider on path p^*, then it must be a collider on path p by construction. Since path p is unblocked with respect to C, i.e. every collider k on the path p is in C or has a descendant in C, it follows that every collider $k \neq l$ on the path p^* is in C or has a descendant in C. Similarly, if node $k \neq l$ is a noncollider on path p^*, then it is a noncollider on path p^* by construction. Since path p is unblocked with respect to C, i.e. none of the noncolliders on path p is in C, it follows that if node $k \neq l$ is a noncollider on path p^*, then $k \notin C$.

Lastly, we need to consider node l. Note that node l is on path p^* by construction. Case 1. Node l is a collider on both path p and p^*. If node l is a collider on path p, then l is in C or has a descendant in C. Therefore, having node l as a collider on path p^* does not block path p^*. Case 2. Node l is a noncollider on both path p and p^*. If node l is a noncollider on path p, then l is not in C. Therefore, having node l as a noncollider on path p^* does not block path p^*. Case 3. Node l is a collider on path p and a noncollider on path p^*. If node l is a collider on path p, then path p must have the following form: $i \cdots * \rightarrow l \leftarrow * \cdots * \rightarrow l \leftarrow * \ldots j$, where $*$ represents that an edge can have an arrow end or a tail end. This implies that path p^* must have the following form $i \cdots * \rightarrow l \leftarrow * \ldots j$, hence l cannot be a noncollider on path p^*. Case 4. Node l is a noncollider on path p and a collider on path p^*. Then path p must have the following form: $i \cdots * \rightarrow l \rightarrow \cdots * \leftarrow l \leftarrow * \ldots j$. This path must contain a collider between the instances of l. Let node c be this collider on path p. Note that it follows that c is a descendant of l. Since node c is a collider on path p and since path p is unblocked with respect to conditioning set C, then it follows

either (i) $c \in C$ or c has a descendant in C. Using this, we now show p^* is unblocked with respect to C. By construction, p^* has the following forms: $i \cdots * \rightarrow l \leftarrow * \ldots j$. Since l is a collider on path p^* and c is a descendant of l, following from (i) and (ii), C includes a descendant of l. Therefore, having node l as a collider on path p^* does not block path p^*. If there is more than one collider between the instances of l in p, this reasoning can be repeated for each collider.

Similarly, we can repeat the above procedure until there are no repeating nodes. Hence it follows we can construct a path p^* without any repeating nodes between (i, j) in $\mathcal{G}$ such that p^* is unblocked with respect to C. $\square$

Lemma 7 shows that in order to establish d-separation and d-connection relations in a graph it is sufficient to consider paths without cycles as specified in Definition 2, but one still has to account for the descendants of colliders on such paths.

Definition 9 (appendage)**.** *Given a path p from i to j defined according to Definition 2, let col_p store the colliders on path p. An appendage of p is a directed path without repeated nodes from a collider $c \in col_p$ to another node $k \in V \setminus \{i, j, c\}$.*

We only need to consider directed paths as appendages because the definition of blocked paths only considers descendants of colliders.

In order to capture d-connections due to conditioning on descendants of a collider, we define the notion of an *extended path*.

Definition 10 (extended path)**.** *Let path p be a path generated according to Definition 2. We define an extended path to be the combination of path p with at least one of p's appendages.*

Consequently, we capture all d-connections by considering paths p without repeating nodes of at most $|V| - 1$ edges and at most $|V| - 2$ colliders. Since the length of an appendage can at

most be $|V| - 3$ edges (the appendage cannot include nodes i, j and c), we can for any pair of variables ensure that the set of extended paths $P_{ij}(\tilde{E})$ between nodes i and j given an edge set $\tilde{E}$ is finite, that the extended paths it contains are of finite length and that it captures all possible d-connections between i and j.

B.2 Enforcing Acyclicity and Causal Sufficiency

Our main formulation presented in Chapter 3 assumes the presence of latent confounders and feedback cycles. However, we can naturally assume causal sufficiency or exclude cycles within our modeling framework. Let $E^s = \{i \leftarrow j, i \rightarrow j,\ \forall i, j \in V : i \neq j\}$ be the set of all possible directed edges under the assumption of causal sufficiency (i.e., no bi-directed edges). Then solving CausalIP(E^s) ensures that unobserved confounders are not allowed.

To exclude cycles, we follow the constraints provided in Cussens (2012) and Jaakkola et al. (2010). These constraints are based on the observation that if cycles are not permitted, any subset of vertices in a graph must contain at least one node that has no parent in that subset. Let $R_i = \{C \mid C \subseteq V \setminus i\}$ be all possible subsets of V that exclude i. Let R_i^k be the k^{th} set in R_i, and let K_i index the sets in R_i. Note that exactly one set in R_i must be the set of parent nodes of i. Accordingly, let w_i^k be a binary decision variable where $w_i^k = 1$ if $R_i^k \in R_i$ is the parent set of node i, and $w_i^k = 0$ otherwise. Next, let E_i^k be the set of all incoming edges to i from nodes in R_i^k, and let ρ_i^k be the number of nodes in R_i^k. Then, we can eliminate cycles by adding the following constraints to CausalIP:

$$w_i^k \leq x_e, \quad e \in E_i^k,\ k \in K_i,\ i \in V, \tag{B.1a}$$

$$w_i^k \geq \sum_{e \in E_i^k} x_e - \sum_{e' \in E \setminus E_i^k} x_{e'} - \rho_i^k + 1, \quad i \in V,\ k \in K_i, \tag{B.1b}$$

$$\sum_{k \in K_i} w_i^k = 1, \quad i \in V, \tag{B.1c}$$

$$\sum_{i \in C} \sum_{\substack{k \in K_i: \\ R_i^k \cap C = \emptyset}} w_i^k \geq 1, \quad C \subseteq V. \tag{B.1d}$$

Constraint (B.1a) ensures R_i^k can only be the parent set of i if the edge $j \to i$ is present for each $j \in R_i^k$. Constraint (B.1b) ensures that if $j \to i$ is present for all $j \in R_i^k$ and if $j \to i$ is not present for all j such that $j \notin R_i^k$, then R_i^k must be the set of parents of node i. Constraint (B.1c) ensures that only one set in R_i can be the parent set of node $i \in V$. Constraint (B.1d) is the directed cycle elimination constraint, which ensures that all subsets of V must contain at least one node who has no parent in that subset.

Our formulation also allows for integrating background knowledge; the presence or absence of specific edges or paths can be easily encoded via constraints on the $\mathbf{x}$ variable. Furthermore, sparsity constraints, such as maximum degree of nodes (see Claassen et al. (2013) for an example) can be easily incorporated into the model. This flexibility may be especially useful in applications where significant domain knowledge is available.

B.3 Characterization of Markov Blankets with Latent Confounders

For DAGs that satisfy the causal Markov and faithfulness conditions, the Markov blanket of target variable T is given by the union of its parents, children, and *spouses*, which are the parents of the children of T (Pearl, 2000) – see Figure B.1(a) for an example. Here we provide a characterization

of Markov blankets that generalizes to DMGs, which allow for both cycles and confounders. We first require the following definition.

Definition 11 (Collider paths)**.** *A path p between nodes $i, j \in V$ is a collider path if every variable on the path p excluding i and j is a collider.*

For example, in Figure B.1(b), the path $T \to n \leftrightarrow o \leftrightarrow p$ is a collider path. This definition also subsumes paths with length 1, so that all children and parents of T also form collider paths with T. Using Definition 11, we can now characterize the Markov blanket for DMGs:

Proposition 6 (Markov blankets for directed mixed graphs)**.** *For a target variable T in a DMG $\mathcal{G}$, the Markov blanket $MB(T)$ is given by the set of all nodes that form a collider path with T.*

Proposition 6 can be shown to be equivalent to the characterization of Markov blankets provided by Pellet and Elisseeff (2008), who focus on an acyclic setting. Our result confirms that this characterization extends to DMGs as well. In the case of DAGs, only parents, children, and spouses of T can form a collider path with T, which makes Proposition 6 specialize to the standard description of Markov blankets for DAGs. Figure B.1(b) provides an illustrative example, where the Markov blanket of T is given by the shaded nodes. The proof of Proposition 6 follows.

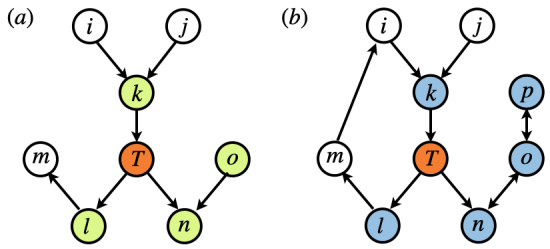


Figure B.1: Example Markov blankets: (a) A DAG where the Markov blanket of T consists of nodes $\boldsymbol{k, l, n}$ and $\boldsymbol{o}$. (b) A DMG where the Markov blanket of T consists of nodes $\boldsymbol{k, l, n, o}$ and $\boldsymbol{p}$.

Proof of Proposition 6. Let $MB(T)$ be a set that includes all nodes that form a collider path with the target variable T. Note that by Remark 1, variables i and j are conditionally independent

(dependent) with respect to C if and only if nodes i and j in the graph are d-separated (d-connected) with respect to C. To prove $MB(T)$ is a Markov blanket of T, it suffices to show that $MB(T)$ is the smallest set that satisfies

$$i \perp T | MB(T), \quad i \in V \setminus \{MB(T), T\}. \tag{B.2}$$

The proof proceeds in two steps. First, we show $MB(T)$ satisfies (B.2). Second, we show that there is no set strictly smaller than $MB(T)$ that satisfies (B.2). **Step 1.** Pick any $i \in V \setminus \{MB(T), T\}$. We consider two cases: If there is no path between i and T, and if there is a path between i and T. First, if there is no path between i and T, then $i \perp T | MB(T)$ holds trivially. Next, suppose there is at least one path p between i and T. Because $i \notin MB(T)$, by construction of $MB(T)$, i is not adjacent to T, and there are no collider paths between i and T. Therefore any path between i and T must include a non-collider. Let p be any such path. Next, to prove $i \perp T | MB(T)$, it remains to show that path p is blocked with respect to $MB(T)$ (Definition 4). Let node j be the non-collider closest to T on path p. To show path p is blocked with respect to $MB(T)$, it suffices to show $j \in MB(T)$ (Definition 3). Let p_{T-j} represent the subpath between nodes T and j on path p. If p_{T-j} is of length one, then j is adjacent to T, or there is a bi-directed edge between j and T. In all three cases, j forms a collider path with T, and thus $j \in MB(T)$. Now suppose the length of p_{T-j} is at least two. Then, because j is the closest non-collider to T, every node between T and j on p_{T-j} must be a collider. Therefore, p_{T-j} is a collider path between T and j, which again implies $j \in MB(T)$. Because j is a non-collider on path p and $j \in MB(T)$, by Definition 3 path p is blocked with respect to $MB(T)$, as desired. **Step 2**. We now prove that there is no set strictly smaller than $MB(T)$ that satisfies (B.2). To do so, we show that any set that satisfies (B.2) must also contain all nodes that form a collider chain with T. Let $MB'(T)$ be a set that satisfies (B.2).

The proof proceeds in two steps: First, we show that all nodes that are adjacent to T must be in $MB'(T)$, and second, that all nodes that form a collider chain with T of length at least 2 must be in $MB'(T)$. **Step 2.1**. Note that for any node i adjacent to T, we have $i \not\perp T|C$ for any conditioning set C. Therefore, because $MB'(T)$ satisfies (B.2), all nodes adjacent to T must be included in $MB'(T)$. **Step 2.2**. We now show that all nodes that form a collider path with T of length at least 2 must be in $MB'(T)$. By Definition 11, it suffices to show that for any node i such that there is a collider path between i and T, $i \in MB'(T)$. Pick any such i, let p be the collider path, and let ℓ be its length. We shall prove the result by induction, using $\ell = 2$ as the base case. Base case. By way of contradiction, suppose $i \notin MB'(T)$. Because $\ell = 2$, there exists a single collider j that is adjacent to T on path p. Because j is adjacent to T, it follows from Step 1 that $j \in MB'(T)$. Because $j \in MB'(T)$, it follows that path p is unblocked with respect to $MB'(T)$ (Definition 3), and thus $i \not\perp T|MB'(T)$ (Definition 4). However, because $i \notin MB'(T)$, $i \not\perp T|MB'(T)$ implies that $MB'(T)$ does not satisfy (B.2) – a contradiction. It follows that $i \in MB'(T)$. Induction step. For the induction hypothesis, assume $i \in MB'(T)$ where p has length less than or equal to ℓ. Next, suppose by way of contradiction that $i \notin MB'(T)$ when p has length $\ell + 1$. Let j be any non-endpoint node on path p. Let p_{T_j} represent the subpath between nodes T and j on path p. Since p is a collider path by assumption, it follows that p_{T_j} is also a collider path with a length less than ℓ. It follows from the induction hypothesis that $j \in MB'(T)$. Let V_p and $\bar{V}_p$ be the sets of colliders and non-colliders on path p. Using the same argument, it is straightforward to show $V_p \subseteq MB'(T)$. Since $\bar{V}_p = \emptyset$, following from p being a collider path, and $V_p \in MB'(T)$, we have $i \not\perp T|MB'(T)$ (Definition 3). However, because $i \notin MB'(T)$, $i \not\perp T|MB'(T)$ implies that $MB'(T)$ does not satisfy (B.2), a contradiction. We conclude $i \in MB'(T)$. We have established that for any $MB'(T)$ that satisfies (B.2), $MB'(T)$ must contain all nodes that form a collider path with T. Therefore, there is no set strictly smaller than $MB(T)$ that satisfies (B.2). $\square$

B.4 Proofs

B.4.1 Proof of Proposition 3

Lemma 8. *Let* $(\mathbf{x}, \mathbf{y})$ *be a solution to the inequalities* (3.4) *for any* $\tilde{E}$*, and let* $\mathcal{G}(\mathbf{x}) = (V, E)$ *be the graph encoded by* $\mathbf{x}$*. Then for each* $p \in \mathcal{P}(\tilde{E})$*,* $y_p = 1$ *if and only if* $p \in \mathcal{P}(E)$*.*

Proof of Lemma 8. Because $\mathcal{G}(\mathbf{x})$ is the graph corresponding to $\mathbf{x}$, for all $e \in \tilde{E}$ we have $x_e = 1$ if and only if $e \in E$. For each $p \in \mathcal{P}(\tilde{E})$, let $E_p \subseteq \tilde{E}$ be the set of edges on path p, and note $\ell_p = |E_p|$. Next, pick any $p \in \mathcal{P}(\tilde{E})$. We first show that $y_p = 1$ implies $p \in \mathcal{P}(E)$. By inequality (3.4b), $|E_p| \leq \sum_{e \in \tilde{E}} \phi_{pe} x_e$. Because $\phi_{pe} = 1$ if and only if $e \in E_p$, we have $\sum_{e \in \tilde{E}} \phi_{pe} x_e = \sum_{e \in E_p} x_e$. Then inequality (3.4b) can be re-written as $|E_p| \leq \sum_{e \in E_p} x_e$. It follows that $x_e = 1$ for all $e \in E_p$, which implies $E_p \subseteq E$. Because $E_p \subseteq E$, it follows that $p \in \mathcal{P}(E)$. Next, we show $p \in \mathcal{P}(E)$ implies $y_p = 1$. By again noting that $\sum_{e \in \tilde{E}} \phi_{pe} x_e = \sum_{e \in E_p} x_e$, we can re-write inequality (3.4a) as $y_p \geq \sum_{e \in E_p} x_e - (|E_p| - 1)$. Because $p \in \mathcal{P}(E)$, and $e \in E$ if and only if $x_e = 1$, we have $x_e = 1$ for all $e \in E_p$. We can then simplify inequality (3.4a) further to $y_p \geq |E_p| - (|E_p| - 1)$, which implies $y_p = 1$. $\square$

Lemma 9. *Suppose Assumption 6 holds. Then there exists a solution* $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ *to* CausalIP(E^c) *such that* $\sum_{i,j \in V} \sum_{n \in N_{ij}} \tilde{z}_{ij}^n = 0$*.*

Proof of Lemma 9. Our approach is to construct a solution $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ such that the following two conditions are satisfied: (i) $\sum_{i,j \in V} \sum_{n \in N_{ij}} \tilde{z}_{ij}^n = 0$, and (ii) $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ is a feasible solution for CausalIP(E^c). The proof proceeds in two steps. First, we construct $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$. Second, we show it satisfies conditions (i) and (ii) above. **Step 1.** Let $\mathcal{G}_T = (V, E_T)$ be the true graph. First, construct $\tilde{\mathbf{x}}$ so that for all $e \in E^c$, $\tilde{x}_e = 1$ if and only if $e \in E_T$. It follows that $\mathcal{G}(\tilde{\mathbf{x}}) = \mathcal{G}_T$. Next, it is straightforward to show using a similar argument to the proof of Lemma 8 that for any fixed

$\mathbf{x}$, there exists a solution to the inequalities (3.4) over $\mathbf{y}$ that satisfies $\mathbf{y} \in \{0,1\}^{|P(E^c)|}$. Let $\tilde{\mathbf{y}}$ be a solution to (3.4) under $\tilde{\mathbf{x}}$. Lastly, let $\tilde{z}_{ij}^n = 0$ for all $n \in N_{ij}$, $i, j \in V$. **Step 2.** Note $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ trivially satisfies condition (i) above. It remains to show that $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ is feasible to CausalIP(E^c). By construction $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies the constraints (3.4) as well as the constraints that $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are all binary-valued. It remains to show that $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies constraints (3.5) and (3.6). To see that $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies (3.5), pick any $n \in N_{ij}^I$, $p \in P_{ij}(E^c)$ and $i, j \in V$. Because $n \in N_{ij}^I$ and Assumption 6 holds, we have $i \perp j | C_{ij}^n$ in graph $\mathcal{G}_T$. Because p is a path from i to j, $i \perp j | C_{ij}^n$ implies that p must be blocked with respect to $C_{ij}^n \in I_{ij}$ (Definition 4). It follows that $\alpha_{ijp}^n = 0$, which implies $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies (3.5). Next, we show $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies (3.6). Pick any $n \in N_{ij}^D$ and $i, j \in V$. Note that by Lemma 8, because $\mathcal{G}(\mathbf{x}) = \mathcal{G}_T$, for all $p \in \mathcal{P}(E^c)$, we have $\tilde{y}_p = 1$ if and only if $p \in \mathcal{P}(E_T)$. It follows that the left hand side of constraint (3.6) can be re-written as $\sum_{p \in P_{ij}(E^c)} \alpha_{ijp}^n \tilde{y}_p = \sum_{p \in P_{ij}(E_T)} \alpha_{ijp}^n$. Because $z_{ij}^n = 0$, it remains to show $\sum_{p \in P_{ij}(E_T)} \alpha_{ijp}^n \geq 1$. Because Assumption 6 holds, $n \in N_{ij}^D$ implies that $i \not\perp j | C_{ij}^n$ in $\mathcal{G}_T$. It follows from Definition 4 that there must exist at least one path between i and j that is unblocked with respect to C_{ij}^n. By definition of α_{ijp}^n, it follows that there exists at least one $p \in P_{ij}(E_T)$ such that $\alpha_{ijp}^n = 1$, which implies $\sum_{p \in P_{ij}(E_T)} \alpha_{ijp}^n \geq 1$, as desired. $\square$

Proof of Proposition 3. It follows immediately from Assumption 6 that $\mathcal{G}_T$ minimizes the objective in (3.3), which implies that any graph $\mathcal{G} \sim \mathcal{G}_T$ also minimizes (3.3). Therefore, under Assumption 6, statement (i) follows directly from (ii). It therefore suffices to show that (ii) holds. Let $(\mathbf{x}^c, \mathbf{y}^c, \mathbf{z}^c)$ be an optimal solution to CausalIP(E^c), where $\mathcal{G}^c := \mathcal{G}(\mathbf{x}^c)$ is the corresponding graph. Our approach will be to show that $\mathcal{G}^c$ satisfies the following two conditions for all pairs (i, j) such that $i \neq j$: (a) for each $C \in D_{ij}$, the nodes i and j are d-connected with respect to C in $\mathcal{G}^c$, and (b) for each $C \in I_{ij}$, nodes i and j are d-separated with respect to C in $\mathcal{G}^c$. Note that

if these two conditions hold, then it follows immediately from Assumption 6 that G^c and $\mathcal{G}_T$ are Markov equivalent. Pick any (i, j), and let it be fixed in the remainder of the proof. We first prove condition (a) holds. Note i and j are d-connected with respect to a conditioning set C if and only if there exists an unblocked path from i to j with respect to C (Definition 4). Suppose by way of contradiction that there exists a conditioning set $C_{ij}^{\bar{n}} \in D_{ij}$ such that all paths between i and j in $\mathcal{G}^c$ are blocked with respect to $C_{ij}^{\bar{n}}$. Then by definition, $\alpha_{ijp}^{\bar{n}} = 0$ for all $p \in P_{ij}(E^c)$. It follows that $\sum_{p \in P_{ij}(E^c)} \alpha_{ijp}^{\bar{n}} y_p^c = 0$. Next, note $\bar{n} \in N_{ij}^D$ because $C_{ij}^{\bar{n}} \in D_{ij}$. Because $\sum_{p \in P_{ij}(E^c)} \alpha_{ijp}^{\bar{n}} y_p^c = 0$, it follows from constraint (3.6) that $z_{ij}^{\bar{n}c} = 1$, and thus $\sum_{i,j \in V} \sum_{n \in N_{ij}} z_{ij}^{nc} > 0$. However, by Lemma 9, there exists a solution $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ to CausalIP(E^c) such that $\sum_{i,j \in V} \sum_{n \in N_{ij}} \tilde{z}_{ij}^{n} = 0$, which yields a contradiction. We conclude that condition (a) holds. We now prove condition (b) holds. Pick any conditioning set $C_{ij}^{\bar{n}} \in I_{ij}$. It suffices to show that all paths between i and j are blocked with respect to $C_{ij}^{\bar{n}}$ in $\mathcal{G}^c$. By definition, $\alpha_{ijp}^{\bar{n}} = 1$ for all paths $p \in P_{ij}(E^c)$ that are unblocked with respect to $C_{ij}^{\bar{n}}$. Therefore, it remains to show $\alpha_{ijp}^{\bar{n}} y_p^c = 0$ for all $p \in P_{ij}(E^c)$. Note $\bar{n} \in N_{ij}^I$ because $C_{ij}^{\bar{n}} \in I_{ij}$. Because $\bar{n} \in N_{ij}^I$, it follows from constraint (3.5) that $\alpha_{ijp}^{\bar{n}} y_p^c \leq z_{ij}^{\bar{n}c}$ for all $p \in P_{ij}(E^c)$. Further, it follows from the proof of condition (a) above that $z_{ij}^{\bar{n}c} = 0$ for all $p \in P_{ij}(E^c)$, which implies $\alpha_{ijp}^{\bar{n}} y_p^c = 0$ for all $p \in P_{ij}(E^c)$. It follows that every path between i and j in $\mathcal{G}^c$ is blocked with respect to $C_{ij}^{\bar{n}}$, as desired. Because conditions (a) and (b) both hold, we conclude $\mathcal{G}^c \sim \mathcal{G}_T$. □

B.4.2 Proof of Lemma 1

Lemma 10. *Suppose $\mathcal{G} \sim \mathcal{G}_T$ and Assumption 6 holds. For each $i, j \in V$, if $\mathcal{G}$ contains an edge between (i, j), then $I_{ij} = \emptyset$.*

Proof of Lemma 10. We show that if $I_{ij} \neq \emptyset$, then $\mathcal{G}$ cannot contain an edge between (i, j). Suppose by way of contradiction that it does. This edge is also a path – call it path p. Next, pick any conditioning set $C \in I_{ij}$. Because p does not contain any colliders or non-colliders, the path

p is unblocked with respect to C (Definition 3). Then by Definition 4, the path p is d-connected with respect to C. Because $\mathcal{G} \in \mathcal{M}$ and the path p is d-connected with respect to C, it follows that $i \not\perp\!\!\!\perp j|C$ (Remark 1). However, because $C \in I_{ij}$, by definition of I_{ij} we have $i \perp\!\!\!\perp j|C$ – a contradiction. The result follows. □

Proof of Lemma 1. We prove the statements in order. $\underline{(i)}$. If (i, j, k) forms a collider chain, then the pairs (i, j) and (j, k) both contain an edge. It follows from Lemma 10 that $I_{ij} = I_{jk} = \emptyset$. We now show that if (i, j, k) form a collider chain, then $j \notin C$ for all $C \in I_{ik}$. Suppose by way of contradiction that (i, j, k) form a collider chain and there exists $C \in I_{ik}$ such that $j \in C$. Because node j is the only collider on the chain, and $j \in C$, the path from i to k is unblocked with respect to C (Definition 3), which implies i and k are d-connected with respect to C (Definition 4). Because the chain is d-connected with respect to C, and $\mathcal{G} \sim \mathcal{G}_T$, by Assumption 6 we have $C \in D_{ik}$. However, this is a contradiction because $C \in I_{ik}$. $\underline{(ii)}$. The result follows by parallel argument to (i), where (i, j, k) form a non-collider chain instead of a collider chain, and $j \notin C$ is used in place of $j \in C$. □

B.4.3 Proof of Proposition 4

Lemma 11. *Suppose $\mathcal{G} \sim \mathcal{G}_T$ and Assumption 6 holds. Let e be an edge between a pair (i, j) in $\mathcal{G}$. If $e \notin \tilde{E}_0$, then there exists $k \in V \setminus \{i, j\}$ such that $\mathcal{G}$ contains an edge between at least one of (i, k) or (j, k).*

Proof of Lemma 11. Because e connects (i, j), $I_{ij} = \emptyset$ by Lemma 10. Because $e \notin \tilde{E}_0$, it follows that there exists $k' \in V \setminus \{i, j\}$ such that at least one of $D_{ik'} \neq \emptyset$ or $D_{jk'} \neq \emptyset$ holds. Suppose $D_{ik'} \neq \emptyset$. By definition of $D_{ik'}$ there exists $C \in D_{ik'}$ such that i and k' are d-connected with respect to k', which implies that $\mathcal{G}$ contains a path from i to k'. Let p denote this path. We now consider

two cases: either the path p contains j or does not contain j. Suppose p contains j, and let p' be the subpath from j to k'. If (j, k') contains an edge, the proof is complete. If not, then the path p' must contain a different node $k \in V \setminus \{i, j\}$ such that (j, k) contains an edge. If the path p does not contain j, then either (i, k') contains an edge, or there exists $k \in V \setminus \{i, j\}$ such that (i, k) contains an edge. Therefore, $D_{ik'} \neq \emptyset$ implies that at least one of (i, k) or (j, k) contains an edge. The proof for the case where $D_{jk'} \neq \emptyset$ follows by parallel argument. □

Lemma 12. *The sets $R(\mathbf{z})$ and $\bar{R}(\mathbf{z})$ constructed in* UpdateEdges *satisfy $R(\mathbf{z}) \cup \bar{R}(\mathbf{z}) \neq \emptyset$.*

Proof of Lemma 12. By Algorithm 2, $R(\mathbf{z}) \neq \emptyset$ if and only if there exists $(i, j, k) \in S$ such that $E_{ijk} \not\subset \tilde{E}$, and $\bar{R}(\mathbf{z}) \neq \emptyset$ if and only if there exists $(i, j, k) \in \bar{S}$ such that $\bar{E}_{ijk} \not\subset \tilde{E}$. Suppose by way of contradiction that $R(\mathbf{z}) = \bar{R}(\mathbf{z}) = \emptyset$. It follows that $E_{ijk} \subset \tilde{E}$ for all $(i, j, k) \in S$ and $\bar{E}_{ijk} \subset \tilde{E}$ for all $(i, j, k) \in \bar{S}$. Next, because UpdateEdges is called, we must have $\tilde{E} \neq E^c$ by Algorithm 1. It follows that there exists an edge $\bar{e} \in E^c$ such that $\bar{e} \notin \tilde{E}$. Because $\bar{e} \notin \tilde{E}$, we have $\bar{e} \notin E_{ijk}$ for all $(i, j, k) \in S$ and $e \notin \bar{E}_{ijk}$ for all $(i, j, k) \in \bar{S}$. Next, let (i, j) be the pair of nodes that edge e connects. Because $e \notin \tilde{E}$ and $\tilde{E}_0 \subseteq \tilde{E}$ by Algorithm 1, we have $e \notin \tilde{E}_0$. It follows from Lemma 11 that there exists a node $k \in V \setminus \{i, j\}$ such that $\mathcal{G}$ contains an edge between at least one of (i, k) or (j, k). Suppose without loss of generality that (j, k) contains an edge. Because (i, j) and (j, k) both contain an edge, the triple (i, j, k) forms a chain. We now have two cases to address: either j is a collider or j is a non-collider. Suppose j is a collider. Then $(i, j, k) \in S$ by Lemma 1. It follows that $e \in E_{ijk}$. Because $E_{ijk} \subset \tilde{E}$ by assumption, we have $e \in \tilde{E}$ – a contradiction. The case where j is a non-collider also yields a contradiction by a similar argument. We conclude that $R(\mathbf{z}) \cup \bar{R}(\mathbf{z}) \neq \emptyset$. □

Lemma 13. NewEdgesIP *is always feasible when called by* UpdateEdges.

Proof of Lemma 13. The formulation NewEdgesIP is called in two cases: $S(\mathbf{z}) \cup S(\mathbf{z}) \neq \emptyset$ and

$S(\mathbf{z}) \cup S(\mathbf{z}) = \emptyset$. The proofs for the second case is identical to the first, where $R(\mathbf{z})$ and $\bar{R}(\mathbf{z})$ is used in place of $S(\mathbf{z})$ and $\bar{S}(\mathbf{z})$. We therefore focus on the case where $S(\mathbf{z}) \cup S(\mathbf{z}) \neq \emptyset$. To prove the result, it suffices to construct a solution $\tilde{\mathbf{w}}$ that satisfies each constraint in NewEdgesIP. Accordingly, let $\tilde{w}_{ij}^t = 1 - \lambda_{ij}^t$ for all $i, j \in V$ and $t \in \{1, 2, 3\}$. With a slight abuse of notation, let e_{ij}^t be the edge corresponding to the variable w_{ij}^t. The proof proceeds in three steps. First, we show $\tilde{\mathbf{w}}$ satisfies constraints (3.16a)–(3.16c); second, that it satisfies (3.17a)–(3.17d); and third, (3.18a)–(3.18c). **Step 1.** We first show that $\tilde{\mathbf{w}}$ satisfies constraints (3.16a)–(3.16c). If $S(\mathbf{z}) = \emptyset$, the result holds trivially. Suppose $S(\mathbf{z}) \neq \emptyset$. Note $\tilde{w}_{ij}^t + \lambda_{ij}^t = 1$ for all $t \in \{1, 2, 3\}$ by construction. It immediately follows that (3.16a) and (3.16b) are satisfied for all $(i, j, k) \in S(\mathbf{z})$. Next we show $\tilde{\mathbf{w}}$ satisfies (3.16c). Pick any $(i, j, k) \in S(\mathbf{z})$. Note $(i, j, k) \in S(\mathbf{z})$ implies there exists $e \in E_{ijk}$ such that $e \notin \tilde{E}$. By definition of E_{ijk}, we must have either $e = e_{ij}^t$ for some $t \in \{1, 3\}$ or $e = e_{jk}^t$ for some $t \in \{2, 3\}$. Suppose $e = e_{ij}^t$ for some $t \in \{1, 3\}$. Because $e \notin \tilde{E}$, at least one of $\lambda_{ij}^1 = 0$ or $\lambda_{ij}^3 = 0$ must hold. By construction of $\tilde{\mathbf{w}}$, it follows that at least one of $\tilde{w}_{ij}^1 = 1$ or $\tilde{w}_{ij}^3 = 1$ must hold. Therefore, $\sum_{t \in \{1,3\}} w_{ij}^t \geq 1$, which implies constraint (3.16c) is satisfied by $\tilde{\mathbf{w}}$. The case where $e = e_{jk}^t$ for some $t \in \{2, 3\}$ follows by parallel argument. **Step 2.** We now show $\tilde{\mathbf{w}}$ satisfies constraints (3.17a)–(3.17d). If $\bar{S}(\mathbf{z}) = \emptyset$, the result holds trivially. Suppose $\bar{S}(\mathbf{z}) \neq \emptyset$. It is straightforward to verify that constraints (3.17a)– (3.17c) are immediately satisfied because $w_{ij}^t + \lambda_{ij}^t = 1$ for all $i, j \in V$ and $t \in \{1, 2, 3\}$. We now show $\tilde{\mathbf{w}}$ satisfies constraint (3.17d). Pick any $(i, j, k) \in \bar{S}(\mathbf{z})$. By definition of $\bar{S}(\mathbf{z})$, $(i, j, k) \in \bar{S}(\mathbf{z})$ implies there exists $e \in e_{ijk}$ such that $e \notin \tilde{E}$. By definition of $\bar{E}_{ijk}$, we must have either $e = e_{ij}^t$ or $e = e_{jk}^t$ for some $t \in \{1, 2, 3\}$. Suppose $e = e_{ij}^t$ for some $t \in \{1, 2, 3\}$. Because $e \notin \tilde{E}$, at least one of $\lambda_{ij}^1 = 0$, $\lambda_{ij}^2 = 0$, or $\lambda_{ij}^3 = 0$ must hold. It follows that at least one of $\tilde{w}_{ij}^1 = 1$, $\tilde{w}_{ij}^2 = 1$, or $\tilde{w}_{ij}^3 = 1$ must hold. Therefore, $\sum_{t \in \{1,2,3\}} w_{ij}^t \geq 1$, which implies constraint (3.17d) is satisfied by $\tilde{\mathbf{w}}$. The case where $e = e_{jk}^t$ for some $t \in \{1, 2, 3\}$ follows by parallel argument. **Step 3.** Constraints (3.18a)–(3.18c) hold because $\tilde{w}_{ij}^t + \lambda_{ij}^t = 1$ by

construction of $\tilde{\mathbf{w}}$, and $\lambda^1_{ij} = \lambda^2_{ji}$ and $\lambda^3_{ij} = \lambda^3_{ji}$ by definition of λ^t_{ij}, for all $i, j \in V$ and $t \in \{1, 2, 3\}$. $\square$

Proof of Proposition 4. By Lemma 13, NewEdgesIP is always feasible when called in UpdateEdges. Because each decision variable w^t_{ij} is binary (and thus bounded), it follows that NewEdgesIP always produces an optimal solution, $\mathbf{w}^*$. Let e^t_{ij} be the edge in E^c corresponding to the variable w^t_{ij}. Then $E_{\text{new}} = \{e^t_{ij} \in E^c | w^{t*}_{ij} = 1\}$. The remainder of the proof proceeds in two steps. First, we show that the set E_{new} constructed in UpdateEdges satisfies $E_{\text{new}} \neq \emptyset$ and $E_{\text{new}} \cap \tilde{E} = \emptyset$. Second, we prove the main result. **Step 1.** We first show $E_{\text{new}} \neq \emptyset$. By definition of E_{new}, it suffices to show there exists $i, j \in V$ and $t \in \{1, 2, 3\}$ such that $w^{t*}_{ij} = 1$. We consider three cases: $S(\mathbf{z}) \neq \emptyset$, $\bar{S}(\mathbf{z}) \neq \emptyset$, and $S(\mathbf{z}) = \bar{S}(\mathbf{z}) = \emptyset$. Note that in the first two cases, UpdateEdges calls NewEdgesIP$(S(\mathbf{z}), \bar{S}(\mathbf{z}))$ to construct E_{new} (Algorithm 2). If $S(\mathbf{z}) \neq \emptyset$, by constraint (3.16c) there exists $(i, j, k) \in S(\mathbf{z})$ such that $\sum_{t\in\{1,3\}} w^{t*}_{ij} + \sum_{t\in\{2,3\}} w^{t*}_{jk} \geq 1$, and the result follows. Similarly, if $\bar{S}(\mathbf{z}) \neq \emptyset$, by constraint (3.17d) there exists $(i, j, k) \in \bar{S}(\mathbf{z})$ such that $\sum_{t\in\{1,2,3\}}(w^{t*}_{ij} + w^{t*}_{jk}) \geq 1$, and the result follows. Lastly, suppose $S(\mathbf{z}) = \bar{S}(\mathbf{z}) = \emptyset$. Then UpdateEdges calls NewEdgesIP$(R(\mathbf{z}), \bar{R}(\mathbf{z}))$ to construct E_{new}. By Lemma 12, we must have either $R(\mathbf{z}) \neq \emptyset$ or $\bar{R}(\mathbf{z}) \neq \emptyset$. If $R(\mathbf{z}) \neq \emptyset$, the result immediately follows from constraint (3.16c); if $R(\mathbf{z}) \neq \emptyset$, the result follows from constraint (3.17d). We conclude $E_{\text{new}} \neq \emptyset$. Next, to see that $E_{\text{new}} \cap \tilde{E} = \emptyset$, note that by constraint (3.18a), $w^{t*}_{ij} = 1$ implies $\lambda^t_{ij} = 0$. Because $\lambda^t_{ij} = 1$ if and only if $e^t_{ij} \in \tilde{E}$ by definition, it follows that $w^{t*}_{ij} = 1$ implies $e^t_{ij} \notin \tilde{E}$. Because $e^t_{ij} \in E_{\text{new}}$ if and only if $w^{t*}_{ij} = 1$, we have $E_{\text{new}} \cap \tilde{E} = \emptyset$, as desired. **Step 2.** We now prove the statement in Proposition 4. Let $\tilde{E}_t$ be the set of edges in iteration t of EdgeGen. We wish to show that there exists $t' < \infty$ such that CausalIP$(\tilde{E}_{t'})$ has an optimal objective value of 0. By Lemma 9, CausalIP(E^c) has an optimal objective value of 0. Therefore, it suffices to show that there exists $t' < \infty$ such that $\tilde{E}_{t'} = E^c$. Note $\tilde{E}_t = \tilde{E}_{t-1} \cup \tilde{E}_{new}$ by Algorithm 2. Further, $E_{\text{new}} \neq \emptyset$ and $E_{\text{new}} \cap \tilde{E} = \emptyset$ by Step 1.

It follows that $\tilde{E}_t \subset \tilde{E}_{t+1}$ for all $t > 0$. Because $\tilde{E}_t \subset \tilde{E}_{t+1}$ and $\tilde{E}_t \subseteq E^c$ for all $t > 0$, there must exist $t' < \infty$ such that $\tilde{E}_{t'} = E^c$. $\square$

B.4.4 Proof of Theorem 5

Proof of Theorem 5. By a parallel argument to the proof of Proposition 3, statement (ii) implies (i) under Assumption 6. Therefore, it remains to prove (ii). For any graph $\mathcal{G} = (V, E)$, define $\mathbf{x}(\mathcal{G})$ such that for all $e \in E^c$, $x_e(\mathcal{G}) = 1$ if and only if $e \in E$. By Proposition 3, $\mathcal{G}^c \sim \mathcal{G}_T$ for any graph $\mathcal{G}^c$ attained at an optimal solution to CausalIP(E^c). Therefore, to show that $\mathcal{G}^* \sim \mathcal{G}_T$, it suffices to construct $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ such that the following two conditions hold: (a) $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies $\bar{\mathbf{x}} = \mathbf{x}(\mathcal{G}^*)$, and (b) $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is an optimal solution to CausalIP(E^c). The proof proceeds in two steps. First, we construct the solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$. Second, we show that it satisfies conditions (a) and (b). **Step 1**. Let $\tilde{E} \subseteq E^c$ be the set of candidate edges at termination of Algorithm 1, and let $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ be the optimal solution to CausalIP$(\tilde{E})$. Construct $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ as follows. Set $\bar{x}_e = \tilde{x}_e$ for $e \in \tilde{E}$; otherwise, set $\bar{x}_e = 0$. Set $\bar{y}_p = \tilde{y}_p$ if $p \in \mathcal{P}(E)$; otherwise, set $\bar{y}_p = 0$. Set $\bar{\mathbf{z}} = \tilde{\mathbf{z}}$. **Step 2**. Note that condition (a) is satisfied by construction. It remains to show that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies condition (b). By Proposition 4, the optimal solution $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ at termination of Algorithm 1 satisfies $\tilde{z}^n_{ij} = 0$ for all $n \in N_{ij}$, $i, j \in V$. It follows that $\bar{z}^n_{ij} = 0$ for all $n \in N_{ij}$, $i, j \in V$. Because $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ attains the optimal objective value of 0, it remains to show that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ is feasible to CausalIP(E^c). Note $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies the binary constraints in CausalIP(E^c) by construction. In the remainder of the proof, we show that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies the constraints (3.4)–(3.6) in CausalIP(E^c), in order. We first show $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies constraints (3.4a) and (3.4b) in CausalIP(E^c). For the paths $p \in \mathcal{P}(\tilde{E})$, the constraints (3.4a) and (3.4b) in CausalIP(E^c) are immediately satisfied, because those constraints also appear in CausalIP$(\tilde{E})$. We now show (3.4a) and (3.4b) are satisfied for $p \in \mathcal{P}(E^c) \setminus \mathcal{P}(\tilde{E})$.

Note that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ trivially satisfies constraint (3.4b) for $p \in P(E^c) \setminus \mathcal{P}(\tilde{E})$, because we set $\bar{y}_p = 0$ for those paths. We now show constraint (3.4a) is satisfied for all $p \in \mathcal{P}(E^c) \setminus \mathcal{P}(\tilde{E})$. Pick any $p \in \mathcal{P}(E^c) \setminus \mathcal{P}(\tilde{E})$. Then for the right hand side of the constraint (3.4a) for p,

$$\begin{aligned}\sum_{e \in E^c} h_{pe}\bar{x}_e - (l_p - 1) &= \sum_{e \in \tilde{E}} h_{pe}\bar{x}_e + \sum_{e \in E^c \setminus \tilde{E}} h_{pe}\bar{x}_e - (l_p - 1) \\ &= \sum_{e \in \tilde{E}} h_{pe}\bar{x}_e - (l_p - 1),\end{aligned}$$

where the first line follows from separating the summation over E^c, and the second line follows because we set $\bar{x}_e = 0$ for all $e \notin \tilde{E}$. Because $\bar{y}_p \geq 0$, it remains to show that $\sum_{e \in \tilde{E}} h_{pe}\bar{x}_e - (l_p - 1) \leq 0$. By definition, $h_{pe} = 1$ if and only if edge e is on path p, which implies $\sum_{e \in E^c} h_{pe} = l_p$. Because $p \notin P_{ij}(\tilde{E})$, there must exist an edge e on path p such that $e \notin \tilde{E}$. It follows that $\sum_{e \in \tilde{E}} h_{pe} < \ell_p$. Because $\bar{x}_e \leq 1$ and ℓ_p is integer-valued, $\sum_{e \in \tilde{E}} h_{pe} < \ell_p$ implies $\sum_{e \in \tilde{E}} h_{pe}\bar{x}_e - (\ell_p - 1) \leq 0$, as desired. Next, we show $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies constraint (3.5) in CausalIP(E^c). Note that for each $n \in N_{ij}^D$ and $i, j \in V$, the left hand side of the constraint is (3.5) is

$$\begin{aligned}\sum_{p \in P_{ij}(E^c)} \alpha_{ijp}^n \bar{y}_p &= \sum_{p \in P_{ij}(\tilde{E})} \alpha_{ijp}^n \bar{y}_p + \sum_{p \in P_{ij}(E^c) \setminus P_{ij}(\tilde{E})} \alpha_{ijp}^n \bar{y}_p, \\ &= \sum_{p \in P_{ij}(\tilde{E})} \alpha_{ijp}^n \bar{y}_p,\end{aligned}$$

where the first line follows by separating the summation over $P_{ij}(E^c)$, and the second line follows because $\bar{y}_p = 0$ for $p \notin P_{ij}(\tilde{E})$ by construction. Therefore, $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies constraint (3.5) in CausalIP(E^c) if and only if

$$\sum_{p \in P_{ij}(\tilde{E})} \alpha_{ijp}^{\bar{n}} \bar{y}_p \leq 1 - \tilde{z}_{ij}^{\bar{n}}, \quad n \in N_{ij}^D, i, j \in V. \tag{B.5}$$

Note that inequality (B.5) is equivalent to constraint (3.5) in CausalIP($\tilde{E}$). Because $\bar{y}_p = y^*_p$ for $p \in P_{ij}(\tilde{E})$, $\bar{\mathbf{z}} = \tilde{\mathbf{z}}$, and $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies constraint (3.5) in CausalIP($\tilde{E}$), inequality (B.5) holds. We conclude $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies constraint (3.5) in CausalIP(E^c). Lastly, we show $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies constraint (3.6) in CausalIP(E^c). Note that this constraint can be separated into the following two subsets of constraints:

$$\alpha^n_{ijp} y_p \leq z^n_{ij}, \quad p \in P_{ij}(E^c) \setminus P_{ij}(\tilde{E}), n \in N^I_{ij}, i, j \in V, \tag{B.6a}$$

$$\alpha^n_{ijp} y_p \leq z^n_{ij}, \quad p \in P_{ij}(\tilde{E}), n \in N^I_{ij}, i, j \in V. \tag{B.6b}$$

It suffices to show $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies both (B.6a) and (B.6b). Note (B.6a) is satisfied by $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ because $\bar{y}_p = 0$ for $p \notin P_{ij}(\tilde{E})$ by construction, and $\bar{z}^n_{ij} \geq 0$ for all $n \in N^I_{ij}$, $i, j \in V$. Next, note that (B.6b) is constraint (3.6) in CausalIP($\tilde{E}$). Because $\bar{y}_p = \tilde{y}_p$ for $p \in P_{ij}(\tilde{E})$, $\bar{\mathbf{z}} = \mathbf{z}^*$, and $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$ satisfies constraint (3.6) in CausalIP($\tilde{E}$), $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}})$ satisfies (B.6b) as well. □

B.4.5 Proof of Theorem 6

Lemma 14. *Let* $(\mathbf{m}^*, \mathbf{v}^*, \boldsymbol{\xi}^*)$ *be an optimal solution to* BlanketIP. *Under Assumption 6, if* $\lambda \in (0, 1/|V|)$, *then* $\xi^*_i = 0$ *for all* $i \in V \setminus \{T\}$.

Proof of Lemma 14. Let $(\mathbf{m}^*, \mathbf{v}^*, \boldsymbol{\xi}^*)$ be an optimal solution to BlanketIP, and suppose by way of contradiction that $\xi^*_i = 1$ for some $i \in V \setminus \{T\}$. Then, trivially, the objective of BlanketIP is at least $1 - \lambda$. We obtain a contradiction by constructing a feasible solution with an objective strictly less than $1 - \lambda$. Construct a solution $(\tilde{\mathbf{m}}, \tilde{\mathbf{v}}, \tilde{\boldsymbol{\xi}})$ where for all $i \in V \setminus \{T\} : \tilde{m}_i = 1$ if and only if $i \in MB(T)$, $v^n_i = 1$ if and only if $C^n_{Ti} = MB(T)$ for all $n \in N_{Ti}$, and $\xi^*_i = 0$. It is straightforward to verify that under Assumption 6, the solution $(\tilde{\mathbf{m}}, \tilde{\mathbf{v}}, \tilde{\boldsymbol{\xi}})$ satisfies constraints (3.21) and (3.23), and

is thus feasible to BlanketIP. Further, by construction, $(\tilde{\mathbf{m}}, \tilde{\mathbf{v}}, \tilde{\boldsymbol{\xi}})$ attains an objective of

$$\sum_{i \in V \setminus \{T\}} ((1-\lambda) \cdot \tilde{\xi}_i + \lambda \cdot \tilde{m}_i) < \frac{1}{|V|} \sum_{i \in V \setminus \{T\}} \tilde{m}_i \leq \frac{1}{|V|}(|V|-1),$$

where the first inequality follows because $\tilde{\xi}_i = 0$ for all $i \in V \setminus \{T\}$ and $\lambda < 1/|V|$, and the second follows because $\sum_{i \in V \setminus \{T\}} \tilde{m}_i \leq |V|-1$. Finally, it follows from $\lambda < 1/|V|$ that $\frac{1}{|V|}(|V|-1) < 1-\lambda$, which contradicts the optimality of $(\mathbf{m}^*, \mathbf{v}^*, \boldsymbol{\xi}^*)$. The result follows. □

Proof of Theorem 6. It follows from Lemma 14 that the optimal objective is $\sum_{i \in V \setminus \{T\}} \lambda \cdot m_i^*$, which is the fewest number of selected nodes that satisfy constraints (3.21)–(3.23). Then by Proposition 6 in Appendix B.3, it suffices to show that every solution to (3.21)–(3.23) such that $\xi_i = 0$, $i \in V \setminus \{T\}$ corresponds to a set $V(T)$ containing all nodes that form a collider chain with T. Let W_1 be the set of nodes that form a collider chain of length 1 with T – in other words, nodes that are adjacent to T. Let W_2 be the set of nodes that form a collider chain of length *at least* 2 with T. The proof proceeds in three steps. First, we prove the following supporting result: If i and T are d-connected with respect to $V(T)$, then $i \in V(T)$. Second, we show $W_1 \subseteq V(T)$ for any feasible $V(T)$, and third, we show $W_2 \subseteq V(T)$ for any feasible $V(T)$. **Step 1.** If i and T are d-connected with respect to $V(T)$, then $i \not\perp\!\!\!\perp T | V(T)$ (by Remark 1). Because $i \not\perp\!\!\!\perp T | V(T)$ and Assumption 6 holds, by definition of I_{Ti} it follows that $V(T) \notin I_{Ti}$. Next, it is straightforward to verify using the definition of θ_{ij}^n and constraints (3.21a) and (3.21b) that $v_i^n = V(T)$ if and only if $C_{Ti}^n = V(T)$ for some conditioning set C_{Ti}^n. Because $V(T) \notin I_{Ti}$, it follows that $v_i^n = 0$ for all $n \in N_{Ti}^I$. Then constraint (3.23) simplifies to $m_i + \sum_{n \in N_{Ti}^I} v_i^n = m_i \geq 1$, which implies $m_i = 1$. By definition of $V(T)$, $m_i = 1$ implies $i \in V(T)$, as desired. **Step 2.** Because i is adjacent to T for each $i \in W_1$ by construction, by definition of I_{Ti} and Assumption 6 we have $I_{Ti} = \emptyset$ for $i \in W_1$. This implies $N_{Ti}^I = \emptyset$ for $i \in W_1$.

It follows that constraint (3.23) simplifies to $m_i + \sum_{n \in N_{Ti}^I} v_i^n = m_i \geq 1$, for all $i \in W_1$. Because $V(T) = \{i \in V | m_i = 1\}$ by definition, we conclude $W_1 \subseteq V(T)$. **Step 3.** We now show $W_2 \subseteq V(T)$. Let W_2^ℓ store all nodes that form collider paths of length less than or equal to ℓ with T, where $\ell \in \{2, \ldots, |V| - 1\}$ (for paths of length $\ell = 1$, the node is adjacent to T, which is addressed in Step 1). Note $W_2^2 \subseteq \ldots \subseteq W_2^{|V|-1} = W_2$. We prove the result by induction, using $\ell = 2$ as the base case. Base case. Pick any node $i \in W_2^2$. It suffices to show $i \in V(T)$. By definition of W_2^2, there is a collider path of length 2 – call it p – from i to T. Let j be the collider on this path. Note that because $\ell = 2$, the collider j must be adjacent to T. Therefore, $j \in V(T)$ by Step 1. Because j is the only collider on path p and $j \in V(T)$, the path p is unblocked with respect to $V(T)$ (Definition 3). Therefore, i and T are d-connected with respect to $V(T)$ (Definition 4). It follows from Step 1 that $i \in V(T)$. Induction step. For the induction hypothesis, assume $W_2^l \subseteq V(T)$ holds. We now show $W_2^{l+1} \subseteq V(T)$ holds, by showing $i \in V(T)$ for any $i \in W_2^{l+1}$. Pick any $i \in W_2^{l+1}$, and let p be the collider path i forms with T. Because p is a collider path of length $\ell + 1$, every non-endpoint node in p forms a collider path of length less than or equal to ℓ with T. Because $W_2^l \subseteq V(T)$ by the induction hypothesis, it follows that every collider on path p is in $V(T)$. Therefore, path p is unblocked with respect to $V(T)$ (Definition 3), which implies i and T are d-connected with respect to $V(T)$. It follows from Step 1 that $i \in V(T)$, and thus $W_2^{l+1} \subseteq V(T)$. $\square$

Bibliography

Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica: Journal of the Econometric Society*, pages 1041–1063, 1990.

Warren P Adams, Richard J Forrester, and Fred W Glover. Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs. *Discrete Optimization*, 1(2):99–120, 2004.

Constantin F Aliferis, Alexander Statnikov, Ioannis Tsamardinos, Subramani Mani, and Xenofon D Koutsoukos. Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. *Journal of Machine Learning Research*, 11(1), 2010.

Michael Alley, Max Biggs, Rim Hariss, Charles Herrmann, Michael Li, and Georgia Perakis. Pricing for heterogeneous products: Analytics for ticket reselling. *Available at SSRN 3360622*, 2019.

Theodore W Anderson. On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, pages 1148–1159, 1962.

Theodore W Anderson, Donald A Darling, et al. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212, 1952.

Joshua D Angrist. Angrist Data Archive. `https://economics.mit.edu/faculty/angrist/data1/data/angkru1991`, 1991. Accessed: 2021-05-01.

Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

Joshua D Angrist and Alan B Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, 15(4):69–85, 2001.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996.

Anil Aswani, Zuo-Jun Shen, and Auyon Siddiq. Inverse optimization with noisy data. *Operations Research*, 66(3):

870–892, 2018.

Anil Aswani, Zuo-Jun Max Shen, and Auyon Siddiq. Data-driven incentive design in the Medicare Shared Savings Program. *Operations Research*, 2019.

Lennart Baardman, Setareh Borjian Boroujeni, Tamar Cohen-Hillel, Kiran Panchamgam, and Georgia Perakis. Detecting customer trends for optimal promotion targeting. *Manufacturing & Service Operations Management*, 2020.

Patrick Bajari, C Lanier Benkard, and Jonathan Levin. Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370, 2007.

Cynthia Barnhart, Ellis L Johnson, George L Nemhauser, Martin WP Savelsbergh, and Pamela H Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46(3):316–329, 1998.

Mark Bartlett and James Cussens. Integer linear programming for the bayesian network structure learning problem. *Artificial Intelligence*, 244:258–271, 2017.

Dimitris Bertsimas and Nathan Kallus. The power and limits of predictive approaches to observational-data-driven optimization. *arXiv preprint arXiv:1605.02347*, 2016.

Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

Dimitris Bertsimas, Vishal Gupta, and Ioannis Ch Paschalidis. Data-driven estimation in equilibrium using inverse optimization. *Mathematical Programming*, 153(2):595–633, 2015.

Peter J Bickel and Kjell A Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, Volumes I-II Package*. Chapman and Hall, 2015.

John Bound, David A Jaeger, and Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

Roger J Bowden and Darrell A Turkington. *Instrumental variables*. Number 8. Cambridge University Press, 1990.

Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

Michael Buckland and Fredric Gey. The relationship between recall and precision. *Journal of the American society for information science*, 45(1):12–19, 1994.

Kasey S Buckles and Daniel M Hungerman. Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics*, 95(3):711–724, 2013.

David Card. The causal effect of education on earnings. *Handbook of Labor Economics*, 3:1801–1863, 1999.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Timothy CY Chan, Taewoo Lee, and Daria Terekhov. Inverse optimization: Closed-form solutions, geometry, and goodness of fit. *Management Science*, 2018.

Tom Claassen, Joris Mooij, and Tom Heskes. Learning sparse causal models is not NP-hard. *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 2013.

William G Cochran. The Chi-squared test of goodness of fit. *The Annals of Mathematical Statistics*, pages 315–345, 1952.

Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, 2012.

William J Conover. A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association*, 67(339):591–596, 1972.

James Cussens. Bayesian network learning with cutting planes. *Proceedings of the Twenty-Eigth Conference on Uncertainty in Artificial Intelligence*, pages 153—-160, 2012.

Joann F de Zegher, Dan A Iancu, and Hau L Lee. Designing contracts and sourcing channels to create shared value. *Manufacturing & Service Operations Management*, 2017.

S Dietze and M Schäuble. On the relationship between fenchel and Lagrange duality for optimization problems in general spaces. *Optimization*, 16(1):7–14, 1985.

Djellel Difallah, Elena Filatova, and Panos Ipeirotis. Demographics and dynamics of mechanical turk workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 135–143, 2018.

Esther Duflo, Rema Hanna, and Stephen P Ryan. Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78, 2012.

Frederick Eberhardt. Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics*, 3(2):81–91, 2017.

Peyman Mohajerin Esfahani, Soroosh Shafieezadeh-Abadeh, Grani A Hanasusanto, and Daniel Kuhn. Data-driven inverse optimization with imperfect information. *Mathematical Programming*, 167(1):191–234, 2018.

Christopher Ferrall and Bruce Shearer. Incentives and transactions costs within the firm: Estimating an agency model using payroll records. *The Review of Economic Studies*, 66(2):309–338, 1999.

Patrick Forré and Joris M Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. *arXiv preprint arXiv:1807.03024*, 2018.

Tian Gao and Qiang Ji. Local causal discovery of direct causes and effects. *Advances in Neural Information Processing Systems*, 28:2512–2520, 2015.

George-Levi Gayle and Robert A Miller. Identifying and testing models of managerial compensation. *The Review of Economic Studies*, 82(3):1074–1118, 2015.

Dan Geiger and Judea Pearl. *On the logic of influence diagrams.* University of California (Los Angeles). Computer Science Department, 1988.

Dan Geiger, Thomas Verma, and Judea Pearl. Identifying independence in bayesian networks. *Networks*, 20(5): 507–534, 1990.

George Georgiadis and Michael Powell. A/B contracts. 2021.

Kay Giesecke, Gui Liberali, Hamid Nazerzadeh, J George Shanthikumar, and Chung Piaw Teo. Call for papers—Management Science—Special issue on data-driven prescriptive analytics. *Management Science*, 64(6): 2972–2972, 2018.

Fred Glover. Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460, 1975.

Sanford J Grossman and Oliver D Hart. An analysis of the principal-agent problem. *Econometrica: Journal of the Econometric Society*, pages 7–45, 1983.

Vishal Gupta, Brian Rongqing Han, Song-Hee Kim, and Hyung Paek. Maximizing intervention effectiveness. *Management Science*, 2020.

Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

Christopher Harris. You're hired! an examination of crowdsourcing incentive models in human resource tasks. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, pages 15–18. Hong Kong, China, 2011.

Chien-Ju Ho, Aleksandrs Slivkins, Siddharth Suri, and Jennifer Wortman Vaughan. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*, pages 419–429, 2015.

Chien-Ju Ho, Aleksandrs Slivkins, and Jennifer Wortman Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research*, 55: 317–359, 2016.

Bengt Holmstrom. Moral hazard and observability. *The Bell Journal of Economics*, pages 74–91, 1979.

John Joseph Horton and Lydia B Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, pages 209–218, 2010.

Antti Hyttinen, Frederick Eberhardt, and Patrik O Hoyer. Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439, 2012.

Antti Hyttinen, Patrik O Hoyer, Frederick Eberhardt, and Matti Jarvisalo. Discovering cyclic causal models with latent variables: A general SAT-based procedure. *arXiv preprint arXiv:1309.6836*, 2013.

Antti Hyttinen, Frederick Eberhardt, and Matti Järvisalo. Constraint-based causal discovery: Conflict resolution with answer set programming. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 340–349, 2014.

Antti Hyttinen, Paul Saikko, Matti Järvisalo, et al. A core-guided approach to learning optimal causal graphs. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*. International Joint Conferences on Artificial Intelligence, 2017.

Guido W Imbens and Paul R Rosenbaum. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126, 2005.

Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on Amazon Mechanical Turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 64–67, 2010.

Tommi Jaakkola, David Sontag, Amir Globerson, and Marina Meila. Learning Bayesian network structure using LP relaxations. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 358–365, 2010.

Nathan Kallus and Angela Zhou. Minimax-optimal policy learning under unobserved confounding. *Management Science*, 2020.

Désiré Kédagni and Ismael Mourifié. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*, 107(3):661–675, 2020.

Arezou Keshavarz, Yang Wang, and Stephen Boyd. Imputing a convex objective function. In *2011 IEEE International Symposium on Intelligent Control*, pages 613–619. IEEE, 2011.

Bert Kestenbaum. Seasonality of birth: two findings from the decennial census. *Social Biology*, 34(3-4):244–248, 1987.

Toru Kitagawa. A test for instrument validity. *Econometrica*, 83(5):2043–2063, 2015.

Simge Kucukyavuz, Ali Shojaie, Hasan Manzour, and Linhchuan Wei. Consistent second-order conic integer programming for learning bayesian networks. *arXiv preprint arXiv:2005.14346*, 2020.

Volodymyr Kuleshov and Doina Precup. Algorithms for multi-armed bandit problems. *arXiv preprint*

arXiv:1402.6028, 2014.

Jean-Jacques Laffont and David Martimort. *The theory of incentives: The principal-agent model*. Princeton University Press, 2009.

Donald KK Lee and Stefanos A Zenios. An evidence-based incentive system for Medicare's End-Stage Renal Disease Program. *Management Science*, 58(6):1092–1105, 2012.

Mette Lise Lousdal. An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, 15(1):1–7, 2018.

Marco E Lubbecke and Jacques Desrosiers. Selected topics in column generation. *Operations Research*, 53(6):1007–1023, 2005.

James Luedtke, Mahdi Namazifar, and Jeff Linderoth. Some results on the strength of relaxations of multilinear functions. *Mathematical Programming*, 136(2):325–351, 2012.

Lyft. Ride Challenges. URL: https://help.lyft.com/hc/enca/articles/360001943867-Ride-Challenges, 2021.

Marloes H Maathuis, Diego Colombo, Markus Kalisch, and Peter Bühlmann. Predicting causal effects in large-scale systems from observational data. *Nature Methods*, 7(4):247–248, 2010.

Sara Magliacane, Tom Claassen, and Joris M Mooij. Ancestral causal inference. *Advances in Neural Information Processing Systems*, 29:4466–4474, 2016.

Hasan Manzour, Simge Küçükyavuz, Hao-Hsiang Wu, and Ali Shojaie. Integer programming for learning directed acyclic graphs from continuous data. *Informs Journal on Optimization*, 3(1):46–73, 2021.

Dimitris Margaritis and Sebastian Thrun. Bayesian network induction via local neighborhoods. In *Advances in Neural Information Processing Systems*, pages 505–511, 2000.

Winter Mason and Duncan J Watts. Financial incentives and the performance of crowds. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 77–85, 2009.

Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, pages 1269–1283, 1990.

Frank J Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

Christopher Meek. Strong completeness and faithfulness in bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 411–418, 1995.

Velibor V Mišić and Georgia Perakis. Data analytics in operations management: A review. *Manufacturing & Service Operations Management*, 22(1):158–169, 2020.

Sanjog Misra and Harikesh S Nair. A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, 9(3):211–257, 2011.

Sanjog Misra, Anne T Coughlan, and Chakravarthi Narasimhan. Salesforce compensation: An analytical and empirical examination of the agency theoretic approach. *Quantitative Marketing and Economics*, 3(1):5–39, 2005.

Ismael Mourifié and Yuanyuan Wan. Testing local average treatment effect assumptions. *Review of Economics and Statistics*, 99(2):305–313, 2017.

Michael P Murray. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20(4):111–132, 2006.

Harry J Paarsch and Bruce Shearer. Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review*, 41(1):59–92, 2000.

Young Woong Park and Diego Klabjan. Bayesian network learning via topological order. *The Journal of Machine Learning Research*, 18(1):3451–3482, 2017.

Judea Pearl. Causality: models, reasoning and inference. *Cambridge, UK: Cambridge University Press*, 2000.

Jean-Philippe Pellet and André Elisseeff. Finding latent causes in causal networks: an efficient approach based on Markov blankets. *Advances in Neural Information Processing Systems*, 2008.

Jose M Pena, Roland Nilsson, Johan Björkegren, and Jesper Tegnér. Towards scalable and data efficient learning of Markov boundaries. *International Journal of Approximate Reasoning*, 45(2):211–232, 2007.

Postmates. How do bonuses and incentives work? URL: https://support.postmates.com/fleet/articles/228603028-article-How-do-bonuses-and-incentives-work-, 2021.

Roy Radner. Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica: Journal of the Econometric Society*, pages 1127–1148, 1981.

Kari Rantanen, Antti Hyttinen, and Matti Järvisalo. Learning optimal causal graphs with exact search. In *International Conference on Probabilistic Graphical Models*, pages 344–355. PMLR, 2018.

Kari Rantanen, Antti Hyttinen, and Matti Järvisalo. Discovering causal graphs with cycles and latent confounders: an exact branch-and-bound approach. *International Journal of Approximate Reasoning*, 117:29–49, 2020.

Thomas Richardson. *Feedback models: Interpretation and discovery.* PhD thesis, Carnegie Mellon, 1996.

Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1): 145–157, 2003.

R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

William P Rogerson. Repeated moral hazard. *Econometrica: Journal of the Econometric Society*, pages 69–76, 1985.

David EM Sappington. Incentives in principal-agent relationships. *Journal of economic Perspectives*, 5(2):45–66, 1991.

Fritz W Scholz and Michael A Stephens. K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

Aaron D Shaw, John J Horton, and Daniel L Chen. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pages 275–284, 2011.

Bruce Shearer. Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, 71(2):513–534, 2004.

Malcolm J Slakter. A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association*, 60(311):854–858, 1965.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.

Nickolay Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281, 1948.

Liam Solus, Yuhao Wang, and Caroline Uhler. Consistency guarantees for greedy permutation-based causal inference algorithms. *arXiv preprint arXiv:1702.03530*, 2017.

Stephen E Spear and Sanjay Srivastava. On repeated moral hazard with discounting. *The Review of Economic Studies*, 54(4):599–617, 1987.

Peter Spirtes. Directed cyclic graphical representations of feedback models. In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*, pages 491–498. Morgan Kaufmann Publishers Inc., 1995.

Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics*, volume 3, pages 1–28. SpringerOpen, 2016.

Peter Spirtes, Clark N Glymour, Richard Scheines, David Heckerman, Christopher Meek, Gregory Cooper, and Thomas Richardson. *Causation, prediction, and search*. MIT press, 2000.

Douglas Staiger and James H Stock. Instrumental variables regression with weak instruments. *Econometrica: journal of the Econometric Society*, pages 557–586, 1997.

Michael A Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737, 1974.

James Stock. Instrumental variables in economics and statistics. *International Encyclopedia of the Social Sciences*, 2002.

James H Stock, Jonathan H Wright, and Motohiro Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529, 2002.

Reiji Teramoto, Chiaki Saito, and Shin-ichi Funahashi. Estimating causal effects with a non-paranormal method for the design of efficient intervention experiments. *BMC Bioinformatics*, 15(1):1–14, 2014.

Sofia Triantafillou and Ioannis Tsamardinos. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *The Journal of Machine Learning Research*, 16(1):2147–2205, 2015.

Sofia Triantafillou, Fattaneh Jabbari, and Greg Cooper. Causal markov boundaries. *arXiv preprint arXiv:2103.07560*, 2021.

Ioannis Tsamardinos, Constantin F Aliferis, and Alexander R Statnikov. Algorithms for large scale Markov blanket discovery. In *FLAIRS Conference*, volume 2, pages 376–380, 2003.

Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, pages 436–463, 2013.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.

François Vanderbeck and Laurence A Wolsey. An exact algorithm for IP column generation. *Operations Research Letters*, 19(4):151–159, 1996.

Marcos Vera-Hernandez. Structural estimation of a principal-agent model: Moral hazard in medical insurance. *RAND Journal of Economics*, pages 670–693, 2003.

Thomas Verma and Judea Pearl. *Equivalence and synthesis of causal models*. UCLA, Computer Science Department, 1991.

Xingyu Wu, Bingbing Jiang, Kui Yu, Huanhuan Chen, et al. Accurate Markov boundary discovery for causal feature selection. *IEEE Transactions on Cybernetics*, 50(12):4983–4996, 2019.

Sandeep Yaramakala and Dimitris Margaritis. Speculative markov blanket discovery for optimal feature selection. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, pages 4–pp. IEEE, 2005.

Ming Yin, Yiling Chen, and Yu-An Sun. The effects of performance-contingent financial incentives in online labor markets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, 2013.

Kui Yu, Lin Liu, Jiuyong Li, and Huanhuan Chen. Mining Markov blankets without causal sufficiency. *IEEE Transactions on Neural Networks and Learning Systems*, 29(12):6333–6347, 2018.

Kui Yu, Lin Liu, and Jiuyong Li. Learning Markov blankets from multiple interventional data sets. *IEEE Transactions*

on Neural Networks and Learning Systems, 31(6):2005–2019, 2019.

Zhalama, Jiji Zhang, Frederick Eberhardt, and Wolfgang Mayer. SAT-based causal discovery under weaker assumptions. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*, 2017.

Jiji Zhang and Peter Spirtes. Strong faithfulness and uniform consistency in causal inference. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, pages 632–639, 2002.