THE ROLE OF UNOBSERVABLES IN STATISTICS AND ECONOMETRICS

Paul Delatte

University of Southern California, *delatte@usc.edu*

Many statistical models are framed in terms of unobservable random variables for which no realizations are ever accessible. These include, for instance, measurement error models, structural econometric models with unobserved heterogeneity, and causal models with potential outcomes. All these models derive from a generic class of problems – inverse problems – whose main features are delineated in this paper. Our discussion helps clarify a number of issues that stem from unobservability and, notably, problems of identification. More importantly, it makes clear that it is the presence of unobservables that confers to these models any explanatory power while, at the same time, severs them largely from empirical support. By contrast, we show that all statistical models where unobservables play no role are completely propped up by the data but are, as a consequence, unable of supporting any explanation. They can be seen as tackling the simpler task of description that consists in assigning a distribution to realizations of it. This principled difference between statistical models helps make clear that unobservables are the fundamental entities that link data to scientific theories.

Keywords: model-based inference; unobservables; inverse problems; identification; specification.

1 Introduction

A distinctive feature of many statistical models used in the physical and social sciences is that they include unobservable random variables. This means that, when these models are framed as parametrized family of measures $\{P^\theta : \theta \in \Theta\}$ where $\theta_0 \in \Theta$ is the true unknown parameter to be recovered, one can never directly sample from P^{θ_0}. In other words, the recovery of θ_0 is necessarily indirect and call for extra information not contained in the data at hand. This way of statistical modeling is extremely common in applications. We illustrate this prevalence in the next examples where we introduce three models from recently published papers in which unobservables feature prominently: a measurement error model in Hu, Schennach, and Shiu (2022), a structural econometric model with unobserved heterogeneity in Moon, Shum, and Weidner (2018), and a causal inference model in Lei and Candès (2021).

Example 1.1 (Hu, Schennach, and Shiu (2022))**.** The paper by Hu, Schennach, and Shiu considers a non-parametric monotonic regression model with measurement error in the regressor. In their

This version: June 18, 2025. The paper supersedes a working paper previously circulated under the title "Description and Explanation". The paper is the result of fruitful discussions mostly held at the University of Southern California. We thank, in particular, Julian Duggan, Hyungsik Roger Moon, and Geert Ridder for very helpful comments.

notation, the model reads as

$$Y = m_0(X^*) + \eta\,,$$
$$X = X^* + \varepsilon\,,$$

where $m_0(\cdot)$ is an unknown function assumed monotonic and Y, X, X^*, η, and ε are real-valued random variables. The variables X^*, η, and ε are assumed unobservable in the sense that no realizations of them can be accessed. The variables X and Y are the only observable ones: all the available data for the problem is $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and is construed as an i.i.d. sample drawn from the joint distribution $P_{X,Y}$ of (X, Y). The objective set by the authors is then to recover the function $m_0(\cdot)$ from the available data. The resulting statistical problem and its dependence on the unobservables will be made clear in Appendix C.1. The interpretation of the model, and notably the interpretation of ε as measurement error and η as a random factor in the decomposition of Y in terms of X^*, will be made explicit in Appendix D.1.

Example 1.2 (Moon, Shum, and Weidner (2018))**.** The paper by Moon, Shum, and Weidner considers a random coefficients logit demand model with interactive fixed effects. In their notation (slightly amended for consistency), the model reads as

$$\delta_{jt} = X_{jt}'\beta^0 + \lambda_j^{0\prime} f_t^0 + e_{jt},$$
$$S_{jt} = \int \pi(\delta_t, X_t, v)\, dG(v; \alpha^0),$$

for all $j = 1, 2, \ldots, J$ and all $t = 1, 2, \ldots, T$ where

$$\pi(\delta_t, X_t, v) = \frac{\exp(\delta_{jt} + X_{jt}'v)}{1 + \sum_{l=1}^{J} \exp(\delta_{lt} + X_{lt}'v)},$$

and G is a known distribution up to α^0. The values α^0, $\beta^0 = (\beta_1^0, \ldots, \beta_K^0)$, $\lambda^0 = ((\lambda_{jr}^0)_{j=1,\ldots,J})_{r=1,\ldots,R}$, and $f^0 = ((f_{tr}^0)_{t=1,\ldots,T})_{r=1,\ldots,R}$ in $\mathbb{R}^L, \mathbb{R}^K, \mathbb{R}^{J\times R}$, and $\mathbb{R}^{T\times R}$ are taken as unknown parameters for the problem. The observables are the random variables $X = (((X_{kjt})_{k=1,\ldots,K})_{j=1,\ldots,J})_{t=1,\ldots,T}$, $S = ((S_{jt})_{j=1,\ldots,J})_{t=1,\ldots,T}$ and $Z = (((Z_{mjt})_{m=1,\ldots,M})_{j=1,\ldots,J})_{t=1,\ldots,T}$ taking values in $\mathbb{R}^{K\times J\times T}$, $\mathbb{R}^{J\times T}$, and $\mathbb{R}^{M\times J\times T}$, respectively. The unobservables are the random variables $((e_{jt})_{j=1,\ldots,J})_{t=1,\ldots,T})$ and $((\delta_{jt})_{j=1,\ldots,J})_{t=1,\ldots,T}$ both taking values in $\mathbb{R}^{J\times T}$. The only available data for the problem is thus the vector $((((x_{kjt})_{k=1,\ldots,K})_{j=1,\ldots,J})_{t=1,\ldots,T}, ((s_{jt})_{j=1,\ldots,J})_{t=1,\ldots,T}, (((z_{mjt})_{m=1,\ldots,M})_{j=1,\ldots,J})_{t=1,\ldots,T})$, which is construed as a sample of size 1 from the joint distribution $P_{X,S,Z}$ of (X, S, Z). The object of interest is chosen by the authors as the vector $(\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$ where $P_{e,X,Z}$ is the joint distribution of e, X, and Z. The resulting statistical problem and its dependence on unobservables will be made explicit in Appendix C.2. The interpretation of the model as capturing aggregate market-level demand will be made clear in Appendix D.2.

Example 1.3 (Lei and Candès (2021))**.** The paper by Lei and Candès considers a standard model for causal inference, known as the Neyman–Rubin potential outcome model, under a binary treatment. For some random variables $Y, Y(1), Y(0), T$ and X taking values in $\mathbb{R}, \mathbb{R}, \mathbb{R}, \{0,1\}$, and $\mathbb{R}^d$, respectively, the model reads as

$$Y = \begin{cases} Y(1) & \text{if } T = 1\,, \\ Y(0) & \text{if } T = 0\,; \end{cases}$$

which rewrites in a more functional form as

$$Y = Y(1)T + Y(0)(1-T).$$

The variables X, Y, and T are assumed observable whereas the variables $Y(1)$ and $Y(0)$ are not. The only available data for the problem is then $(x_i, y_i, t_i)_{i=1,2,\dots,n}$ and is construed as an i.i.d. sample drawn from the joint distribution $P_{X,Y,T}$ of (X, Y, T). The authors are primarily interested in the random variable $\tau = Y(1) - Y(0)$ which they refer to as the individual treatment effect. The resulting statistical problem and its dependence on unobservables will be made clear in Appendix C.3. The interpretation of the model as causal, and in particular the interpretation of the unobservables $Y(1)$ and $Y(0)$ as potential outcomes, will be made explicit in Appendix D.3.

The natural question that follows is: why would a statistician or an econometrician tie one's hands by working with a probability distribution that is impossible to sample from? The answer to that question is the object of this paper. Our main claim is that unobservables are what allow practitioners to link data to a field-specific theory in what forms, in fine, an empirically-supported explanation. They are the necessary bridge that allows scientists to leverage the full statistical apparatus of inference to bring data into a theoretical model. In other words, without unobservables, no empirically-supported explanation. However, this comes at a cost: the additional information that propels the data into an explanatory model is severed from empirical support. This is completely and intuitively captured by the nature of unobservables as entities that cannot be observed. This also implies that unobservables, and the stochastic model they sustain, come after an a priori theoretical frame that confers to them a specific field-dependent meaning – we call this the theory-ladenness of unobservables. It follows that statisticians and econometricians do not so much tie their hands when they work with unobservables and probability distributions they cannot directly sample from, but open data and statistical tools to a wide range of new problems (as demonstrated by the examples above). The cost to pay is that of an unavoidable "leap of faith" from what can be observed to an underlying explanation that cannot be completely supported by empirical means. To show this as clearly as possible, we isolate a simple structure that sustains all statistical models with unobservables. This structure helps formalize the notion of inverse problems in statistics and simplify a number of fundamental questions traditionally faced by practitioners when working with unobservables – and notably the problem of identification.

The widespread use of statistical models with unobservables then brings forth a natural question:

Are all statistical models of this form? It is hard not to answer an emphatic "no" to that question by directly considering one of the simplest inferential problems – non-parametric mean estimation using a set of observations. If this question has an easy answer, it is less clear how to interpret all the statistical models that only include observable random variables, especially in light of our claim that explanations suppose theory-laden unobservables. To make progress on this problem, we exhibit a very simple structure that concisely captures the task a statistician performs when using such statistical models. The idea is that, when limiting oneself to observable quantities only, the sole task left is to assign a plausible distribution for the random phenomenon from which the observations were obtained. This simple act of associating a distribution to its realizations delineates a descriptive task that does not involve any "leap of faith": any a priori assumption about the distribution to facilitate inference can always be confronted with the data. The cost of this streamlined process is that inference cannot go beyond the observables: what is obtained is a description whose interpretation directly derives from the meaning assigned to the observables. While this descriptive task could directly start from a parametrized statistical model, we believe that there is value in exhibiting a structure that stands behind any parametrization. We show, in particular, that taking the observable distributions as logically anterior to any parametrized sets helps clarify a number of issues in model-based inferential problems that only include observables.

The rest of the paper is constructed as follows. We start in Section 2 by clarifying how our discussion fits into the broader statistical and econometric literature. We then tackle statistical models that only feature observables in Section 3 and finally move to the core of the paper in Section 4 where we handle statistical models with unobservables. We should revisit the examples introduced above in Appendix C and Appendix D. For the sake of exposition, we will also introduce a set of elementary statistical problems whose purpose is to contrast the different modeling approaches. These examples are collected in Appendix A and Appendix B.

2 Literature review

The majority of statistical models considered in statistical textbooks only include observable random variables. This is true independently of the era, school, and inferential paradigm – see, e.g., Ferguson (1967), Berger (1985), Lehmann and Casella (1998), Robert (2007), Shao (2008) for a small sample. In these textbooks, statistical problems are introduced through a parametrized family of probability measures $\{P^\theta : \theta \in \Theta\}$ with a true unknown parameter θ_0 to be recovered and the hypothesis that one can directly sample from P^{θ_0} finitely many times. It is only for more complex problems that an uneasy confrontation with unobservable variables can be observed – this confrontation is never formalized and never clearly interpreted. The first problems that generate such a confrontation are usually regression models and, more rarely, missing data models.

The state of affair is different in econometrics where models with unobservable random variables occupy a central position. This prevalence seems mostly due to the fact that many statistical problems considered in the field have for origin a preliminary economic model whose main components are theoretical and not directly measurable. However, a full investigation of the difference

between statistical models without unobservables and statistical with unobservables has never been carried out from first principles. The field has favored solving more practical questions directly implied by unobservability. Indeed, the main challenge for econometricians has been to find ways to circumvent unobservables and recover classical statistical problems so that the standard tools, language, and techniques of statistical inference could be leveraged without modification. This has led the field to focus on the problem of identification, paving the way for a broad and general solution stated in the traditional language of statistics – see, for instance, Matzkin (2007) for an authoritative presentation where problems with unobservables are formalized through parametrized families for the observable distributions and identification is traditionally defined as a property of the parametrization.

These formalization exercises have been intertwined with the development of a subfield of statistics and econometrics whose focus has been the study of inverse problems – see, e.g., Cavalier (2011), Florens and Simoni (2017), Schennach (2020). Inverse problems in statistics have for origin inverse problems in physics for which the final deterministic states are observed with noise: the objective is thus to find $f \in F$ such that $A(f) = g$ where A is a known operator and g is known up to some stochastic disturbance[1]. This origin, however, does not capture the diversity of inverse problems considered in statistics and econometrics that are, in the words of Cavalier (2011), all "[t]hese are problems where we have indirect observations of an object [...] that we want to reconstruct". We show in this paper that it is possible to delineate a simple structure that covers all these problems, independently of their connection to a deterministic inverse problem. This structure is relatively simple and originates from constraining a coupling between the observables' distribution and the unobservables' distribution. It naturally bears resemblance to the construction in Matzkin (2007) – the main difference comes from re-ordering the construction by taking the joint distribution of the observables and the unobservables as a starting point. This makes clear, in particular, that the parametrization of the problem is unnecessary. More importantly, this allows us to simplify the problem of identification and unify all problems with unobservables as inverse problems. It also makes clear that the idea of inversion in statistics does not come from inverting an operator but from "undoing" the effects of unobservable processes – as clearly perceived in Giné and Nickl (2021) pp. 3-4 or in Schennach (2020) p. 489. This "undoing" is nothing more than solving the identification problem.

The goal of this paper is not simply to formalize inverse problems but to make clear the trade-off that unobservables entail in statistical models. This question has been partly ignored in the literature as statisticians and econometricians vied for the safe harbor of statistical models under complete observability once identification was resolved. Even the best accounts on unobservables in the field – Matzkin (2007) or Schennach (2020) – do not make explicit what is gained and what is lost by assuming more than observables. To make progress on this front, we are invited to go to the other side and interpret these models that only include observable random variables. This

[1]This is originally the approach that has been used by physical scientists when connecting physical inverse problems to statistical problems – see, for instance, Evans and Stark (2002), Tarantola (2005), Aster, Borchers, and Thurber (2018) going as far back as Sudakovand and Khalfin (1964).

question has been partly passed over due to the conceptual simplicity of these models: their nature and interpretation derive directly from the capacity to sample from a certain distribution. We make this explicit by introducing what we call forward problems where the descriptive nature of the task is contrasted with that in inverse problems. A different attempt at formalizing similar ideas can be found in Evans and Stark (2002) but the authors only ever define forward problems in statistics in terms of parametrized statistical models (hence begging the question as to the origin of such models). The forward structure we suggest in this paper makes a case for the logical precedence of measures as complete descriptions for observations before any attempt at parametrization.

More generally, we grapple in this paper with the existence of different tasks within statistics. Naturally, we are not the first one to face the issue. The idea of a fundamental difference between statistical models lays latent in the diversity of statistical applications. Several papers have tried to explicitly characterize this difference and offer support for one or another position. The most influential contribution in this direction is without contest the "two cultures" paper of Leo Breiman (Breiman (2001)). The author identifies two tasks within statistics, "prediction" and "information [extraction]", and two criteria to judge statistical results, "prediction accuracy" and "interpretability", and then compares the "data modeling culture" and the "algorithmic modeling culture" to serve these objectives. Breiman, however, remains mostly silent about the meaning of "information [extraction]" and "interpretability" (due to its legitimate attraction to prediction as queen and servant of statistics). Our paper can be understood as showing that "information [extraction]" takes at least two distinct forms, description and explanation, and that the difference between these two tasks stem from the role played by unobservables. As far as we know, we are the first one to confront so directly the question of the nature of "information [extraction]" within statistics. This discussion naturally fits in a larger epistemological literature where the distinction has been amply considered – see, for instance, Woodward and Ross (2021) and the references therein – in particular, Van Fraassen (1980).

Finally, this paper contributes to a larger literature on the foundations of statistics, the nature of statistical models, and the role of statistics in the sciences that has accompanied the field since its inception. These foundational discussions – Fisher (1922, 1925, 1935), Neyman (1952, 1976, 1977), Box (1976), Lehmann (1990), Cox (1990) to only name a few – have not only led to a number of famed aphorisms but have directly oriented statistical practices in fundamental ways. Some particular traits of models with unobservables have motivated a number of discussions closer to ours – see, notably, Mayo (1997) and Spanos and Mayo (2015) where the need to differentiate a "statistical model" from a "substantive model" and the risk of "foisting [a] substantive model on the data" are clearly identified. Our discussion leads to sensibly similar advisory remarks. We believe, however, that the role that unobservables play in statistical models deserves more attention. The problem is not simply that of recognizing the difference between an unidentified model with unobservables and a probability distribution one can sample from. It is, above all, to recognize what one gains when working with unobservables – an empirically-supported explanation – and what one has to pay for that – the impossibility to completely falsify the explanation. While there is a room to improve on the explanation by looking at the implied probability distribution on the

observables, there is a fundamental limit in this exercise (beyond the need of strong parametric assumptions and the restriction to simple problems) that comes from the nature of observables itself. Thereafter, our goal is not so much to devise techniques to test when one is "foisting [a] substantive model on the data", but to make clear the impossibility of doing so. This impossibility is not in vain as it is the minimal price to pay to bridge data and scientific theories. Our discussion also suggests that the best way to make sure one is not "foisting [a] substantive model on the data" is to pay full attention to the unobservables (and the probabilistic assumptions one put on them) in light of a preliminary field-dependent theoretical bedrock. This focus is notably motivated by the (re)emergence of causal inference as a prominent statistical problem where "empirical adequacy" is a very limited guide due to the central position of unobservables in this case. This problem and its successes, where most of the assumptions come from a priori discussions on the unobservables' inherent properties without empirical support, partly challenges the anti-realist picture of Van Fraassen (1980) from within statistics. We will leave, however, a full exploration of these philosophical consequences to future research.

3 Description: an upper bound for observables

Before exploring the role of unobservables in depth, we make a preliminary step and first confront these statistical models where unobservables are absent. The purpose of this detour is to prepare the ground for the next section where the nature of unobservables will be easier to pin down by contrast. The task is, however, of interest in itself: statistical models that only include observables come first and are widespread in applications but are often left without a clear interpretation. The following discussion helps make clear what complete observability implies.

All statistical models that only include observable random variables have the same irreducible structure: the available data $x_1, x_2, \ldots, x_n$ is construed as the realizations of some random variables $X_1, X_2, \ldots, X_n$ that are assumed independent and identically distributed (i.i.d.) according to an unknown probability measure P_X. It is the assumption that one has access to the realizations $x_1, x_2, \ldots, x_n$ that make the random variables $X_1, X_2, \ldots, X_n$ observable. In other words, complete observability means that one can directly sample from some distribution P_X finitely many times.

Remark 3.1. This formulation includes cases with dependence and distributional-shift by taking $n = 1$ and X_1 a random vector whose m-components are sampled according to the joint distribution P_X. Because of this, we should directly work under a general i.i.d. formulation where dependent samples are construed as draw of size 1 from a given multivariate distribution.

While complete observability makes inference a concise problem, it comes at a cost: inference cannot go beyond what is observable. Indeed, given the structure above, there is no other statistical problem than to gain information on P_X from the sample $x_1, x_2, \ldots, x_n$ where P_X has a fixed interpretation as the distribution of some observable phenomenon. In predicating a distribution to some observations, the modeler can be seen as providing a complete description of the random phenomenon that generated them. In other words, when all the variables making up a statistical

problem are observable, the most one can do is to find their distribution: there is nowhere to go beyond what is observable. The advantage of this observational constraint is that the task is completely supported by the data $x_1, x_2, \ldots, x_n$ at hand in the sense that any preliminary assumption on P_X, as well as the sampling assumption, can be empirically falsified by statistical means – we call this empirical closure. The outcome of this procedure is an empirically-closed description that consists in the recovery of P_X from $x_1, x_2, \ldots, x_n$ whose meaning completely derives from the observability of the random phenomenon common to $X_1, X_2, \ldots, X_n$.

To make this point as clearly as possible, we err on the side of caution and break statistical models under complete observability down into digestible bits. We do so because models that only include observables are traditionally introduced as parametrized families of probability measures $\{P^\theta : \theta \in \Theta\}$ where it is assumed that one can sample directly from the true unknown distribution P^{θ_0} finitely many times to recover $\theta_0 \in \Theta$. The risk is that the parametrization is not logically preliminary and may erroneously suggest that, even under complete observability, one may do more than recover the observable distribution P^{θ_0}, that is, obtain more than a simple description. By remembering that an indexed family of probability measures $\{P^\theta : \theta \in \Theta\}$ always supposes the existence of an unindexed family of measures $\mathscr{P}$, an index set Θ, and a surjective function $I\colon \Theta \to \mathscr{P}$ called index map, it is possible to decompose statistical models under complete observability as the successive choice of:

1. a set of probability measures $\mathscr{P}$ that is supposed to contain P_X: this delineates the range of possible descriptions for the phenomenon under study;

2. a well-defined function $f\colon \mathscr{P} \to E$ where E is an arbitrary set: this restricts the quantity and type of information that should be considered for any description.

The resulting statistical task is then the recovery of $f(P_X)$ from the data $x_1, x_2, \ldots, x_n$ which is construed as an i.i.d. sample from P_X. For ease of reference, we collect this decomposition in the following definition where the terminology follows from Remark 3.3 below.

Definition 3.1 (Forward Problem)**.** A forward problem for a distribution P_X, which can be sampled from, is a pair $(\mathscr{P}, f)$ where:

1. $\mathscr{P}$ is a set of probability measures such that $P_X \in \mathscr{P}$, and;
2. $f\colon \mathscr{P} \to E$ is a function for some arbitrary set E.

The decomposition of statistical problems under complete observability through forward problems helps make clear their descriptive limit: the first-best solution is always the recovery of P_X which would completely describe the observable random phenomenon X; because recovering P_X may be statistically challenging, it is common to consider a sub-problem of lesser difficulty by focusing on $f(P_X)$; the recovery of $f(P_X)$ then provides a partial description of the phenomenon X. The choice of $(\mathscr{P}, f)$ thus always captures a certain trade-off between the difficulty of the statistical problem and the completeness of the intended description for X.

We illustrate in Appendix A the applicability of this decomposition by breaking down five standard statistical problems that only include observables. These examples further demonstrate

the descriptive upper limit of statistical problems under complete observability. They also support the precedence of measures as descriptions for observations before any attempt at parametrization. We generalize these points in the following remarks. In particular, we show that:

- a forward problem always induces a parametrized statistical model – see Remark 3.2;
- identification is not a problem under complete observability – see Remark 3.3;
- the condition $P_X \in \mathscr{P}$ is always testable yielding empirical closure – see Remark 3.4;
- the decomposition of statistical models for observables as forward problems offers a simple way to order these models – see Remark 3.5;
- sampling is an important modeling choice that always bears on the descriptive scope of the resulting forward problem – see Remark 3.6.

Remark 3.2. A forward problem $(\mathscr{P}, f)$ for P_X indirectly induces a statistical model for P_X in the sense of a parametrized family of measures containing P_X. It is obtained by considering $\{P^\theta \in \mathscr{P} : \theta \in \Theta\}$ where $\Theta = f(\mathscr{P}) = E$ and $\theta_0 = f(P_X)$. It is important to note, however, that the parametrization is not necessarily a valid parametrization for the original set $\mathscr{P}$ since the induced index relation $\theta = f(P) \mapsto P$ is not necessarily a valid mapping from Θ to $\mathscr{P}$ (as some image may have several preimages). It is the case only when f is a bijective function (which is not guaranteed for many problems of interest). When f is not bijective, the parametrization implies the existence of an unindexed set $\mathscr{P}'$ of measures strictly contained in $\mathscr{P}$ for which the parametrization holds. It hence follows that $\{P^\theta \in \mathscr{P} : \theta \in \Theta\}$ may not be exactly the same structure as $(\mathscr{P}, f)$. However, the difference is mild since the statistical task of gaining knowledge on $f(P_X) = \theta_0$ remains identical. The main issue is different and follows from the fact that, in absence of a forward problem, it is unclear how to come up with well-behaved parametrized families like $\{P^\theta \in \mathscr{P}' : \theta \in \Theta\}$ for some potential set $\mathscr{P}'$ beyond simple problems with natural bijective parametrizations. This suggests that, practically and logically, parametrized statistical models under complete observability are always the result of some forward problems.

Remark 3.3. The notion of identification can be defined for indexed statistical models where the object of inference is the indexing parameter: namely, an indexed statistical model $\{P^\theta : \theta \in \Theta\}$ with inferential object of interest θ is said to be identifiable if the map $\theta \mapsto P^\theta$ is injective. We show that an inferential problem with explicit forward structure $(\mathscr{P}, f)$ never leads to problems of identification. First, if $\mathscr{P}$ is not indexed, there is no problem of identification. If $\mathscr{P}$ is indexed, the index either plays no role or is selected as the object of inference through f, but if it is selected through the choice of f, it is de facto identified as the map $\theta \mapsto P^\theta$ is injective due to the fact that f has to be a well-defined function. This result can be alternatively obtained by considering the induced statistical model $\{P^\theta \in \mathscr{P} : \theta \in \Theta\}$ where $\Theta = f(\mathscr{P}) = E$ with implied unindexed set $\mathscr{P}' \subseteq \mathscr{P}$: the index map $I\colon \Theta \to \mathscr{P}', \theta = f(P) \mapsto P$ is always injective since f is a function and $\Theta = f(\mathscr{P}) = f(\mathscr{P}')$. The absence of identification issues in forward problems is the direct manifestation that statistical models with observables are empirically closed.

Remark 3.4. The problem of specification in statistical models under complete observability corresponds exactly to the choice of the set $\mathscr{P}$ in some forward problem $(\mathscr{P}, f)$. Indeed, a statistical model for P_X is well-specified if the condition $P_X \in \mathscr{P}$ is satisfied. The correct specification of a model entirely captures the validity of the resulting descriptions in the sense that uncertainty quantification around an estimate of $f(P_X)$ can then be interpreted as measuring the plausibility of this description. Due to complete observability, specification is a problem that is, in principle at least, always possible to empirically support – this is what we call empirical closure. One practical solution to ensure specification is to split the observations into two sets where the specification's correctness is assessed on the first part and inference is performed on the second part. Such a procedure presents a number of technical challenges from a statistical viewpoint, but it remains feasible, at least in principle, in all models that only include observables.

Remark 3.5. The decomposition of statistical models under complete observability into forward problems leads to a simple way to order and compare such models. We introduce an equivalence relation and an ordering, which are illustrated through the four first examples of Appendix A.

- Two forward problems $(\mathscr{P}, f)$ and $(\mathscr{P}', f')$ are said to be in the same class if $\mathscr{P} = \mathscr{P}'$. Problems in the same class thus consider the same possible distributions for the observables. They only differ in the parts of the observables' distribution they aim to recover. This is captured by the possible difference between f and f'. All problems in the same class would be solved if the unknown distribution of the observables could be perfectly recovered. The difficulty of inference for problems in the same class then depends exclusively on the complexity of $f(\mathscr{P})$ and $f'(\mathscr{P})$. This is illustrated by the problems of linear regression and non-parametric regressions of Examples A.4 and A.3: the two problems are seen to belong to the same class where $\mathscr{P}$ is the set of distributions on $\mathbb{R}^2$ with finite second moments.

- A forward problem $(\mathscr{P}', f')$ is said to be a sub-problem of another forward problem $(\mathscr{P}, f)$ if $\mathscr{P}' \subset \mathscr{P}$ and if f' is the restriction of f to $\mathscr{P}'$. In other words, a sub-problem considers a simpler inferential problem by restricting the set of possible distributions to be considered. The difficulty of inference is then directly captured by the complexity of the sets $\mathscr{P}$ and $\mathscr{P}'$. For instance, normal mean estimation as introduced in Example A.2 is directly seen to be a sub-problem of non-parametric mean estimation as introduced in Example A.1.

Remark 3.6. Forward problems were introduced in Definition 3.1 under an i.i.d framework with dependence and distributional-shift handled as a special case – see Remark 3.1. This simplification should not hide the fact that sampling is a modeling assumption that bears directly on the descriptive scope of the resulting forward problems. To make this explicit, it is always possible to redefine forward problems for the joint distribution $P_{X_1,\dots,X_n}$ of the full sample $(X_1, X_2, \dots, X_n)$ rather than for an hypothesized identical marginal P_X. In this case, the available data $(x_1, x_2, \dots, x_n)$ is always construed as a sample of size 1 from $P_{X_1,\dots,X_n}$. An i.i.d. assumption can then be seen as the most restrictive hypothesis from the descriptive standpoint: it implies a joint distribution for the full sample that is the most invariant possible as it simply corresponds to the product measure

for a single identical marginal distribution. Any relaxation of the i.i.d. assumption thus extends the range of possible descriptions, even if not explicitly clear from Definition 3.1. The importance of the sampling choice as a descriptive constraint is discussed for the examples of Appendix A in Remark A.3.

4 Explanation: the weight of unobservables

The previous section made clear that, under complete observability, any statistical task is descriptive in nature: one cannot get more than the distribution of the random variables from which one has collected observations. Epistemologically, however, one may want to get more from statistics: in particular, one may want to use observations to recover something that is not directly observable. This is exactly what is asked for in an explanation where observable processes are made sense of by way of an underlying structure removed (both logically and practically) from observability in what constitutes, in fine, a theoretical model or a scientific theory. Not surprisingly, the statistical solution to this problem is to introduce unobservable random variables alongside observable ones in a model that mirrors the theoretical one. It is the objective of this section to make sense of this practice. Our discussion should make clear, in particular, that unobservables are the fundamental entities that allow statisticians and econometricians to step beyond descriptions to output empirically-supported explanations. However, because unobservables are, by nature, severed from empirical support, we should see that these explanations cannot be entirely supported by the data but always calls for a certain "leap of faith". This trade-off will be illustrated at length in Appendices C and D through the examples of the Introduction.

4.1 Unobservables, inverse problems, and identification

Statistical models with unobservables have a different origin and structure than models under complete observability. In particular, the forward problem isolated in the previous section where a family of measures for the observables is chosen first is not relevant anymore.

Nonetheless, it is still possible to delineate a common structure for models with unobservables that is intimately tied to the task they support. All the available data $x_1, x_2, \ldots, x_n$ is still construed as realizations of independent random variables $X_1, X_2, \ldots, X_n$ with identical distribution P_X. However, P_X is now understood as the marginal of a joint distribution $P_{X,U}$ that cannot be sampled from. This embedding follows from the fact that one hopes to recover an object $g(P_{X,U})$ that depends not simply on P_X but on the joint system $P_{X,U}$. The other marginal P_U is construed as the identical distribution of some independent random variables $U_1, U_2, \ldots, U_n$ for which we never have realizations – the so-called unobservables.

Remark 4.1. This formulation includes cases with dependence and distributional-shift by taking $n = 1$ with X_1 and U_1 m-random vectors whose $2m$-components are sampled according to the joint distribution $P_{X,U}$. Because of this, we should directly work under a general i.i.d. formulation where dependent samples are construed as draw of size 1 from a given multivariate distribution.

The common denominator of models with unobservables is thus the recovery of an object that is only indirectly observable. The resulting statistical exercise is made feasible because of the introduction of unobservable random variables and a joint distribution for them and the observables. Before considering the origin of this unobservable system, we first collect the construction in the following definition under what we call an inverse problem.

Definition 4.1 (Inverse Problem)**.** An inverse problem for a distribution P_X, which can be sampled from, is a pair (Γ, g) where:

1. Γ is a coupling for P_X and some other distribution P_U that cannot be sampled from;
2. $g : \Gamma \mapsto \theta_0 = g(\Gamma)$ is a function that defines an object $\theta_0 = g(\Gamma)$ to be recovered from Γ.

The coupling Γ in Definition 4.1 is alternatively denoted $P_{X,U}$. It is always constrained by a set of preliminary conditions $\mathcal{C}(\Gamma)$ that is implicitly part of the inverse problem (Γ, g). These constrains (almost) always stem from the specification of a functional equation between $X \sim P_X$ and $U \sim P_U$ in what constitutes a stochastic model that we should denote $\text{SM}(X, U)$. This stochastic model is often characterized by a number of additional distributional assumptions on the pair (X, U) that bear equivalently on the coupling $\Gamma = P_{X,U}$ and that should be included in $\mathcal{C}(\Gamma)$. The resulting set of conditions $\mathcal{C}(\Gamma)$ then works as a defining constraint on the coupling Γ from which an inverse problem derives. We will argue in Section 4.2 that the constraints $\mathcal{C}(\Gamma)$ always originates from a preliminary field-specific theory. This theoretical bedrock not only confers to the stochastic model $\text{SM}(X, U)$ and the variables (X, U) field-specific interpretations, but also gives a way for the modeler to weigh distributional assumptions on (X, U) beyond what is observable. This will be illustrated at length for the examples of the Introduction in Appendix D.

The difference between forward problems and inverse problems is explicit from the respective Definitions 3.1 and 4.1. It is seen, in particular, that inverse problems are characterized by the presence of variables whose distributions cannot be sampled from – the unobservables. This creates a fundamental problem for inverse problems, partly hidden in their definition but from which the name derives, that the object of interest $g(\Gamma)$ cannot be directly recovered from the available data $x_1, x_2, \ldots, x_n$ since it depends on a distribution P_U that cannot be sampled from. For the problem to be solvable, an additional step may be needed – the identification step. That is, additional a priori assumptions on the coupling Γ may be required so as to guarantee the existence of a known function G such that $g(\Gamma) = G(P_X)$. This step is what allows to reconstruct the object $g(\Gamma)$ of interest from indirect observations $x_1, x_2, \ldots, x_n$ of it. Since the reconstruction amounts to "undoing" the effect of the unobservables on the object of interest, it is also whay justifies us calling these problems inverse – see Remark 4.3 for additional clarifications on the terminology.

Definition 4.2 (Identification)**.** An inverse problem (Γ, g) for P_X is said to be identified if there exists a function G such that $g(\Gamma) = G(P_X)$. The additional assumptions on Γ needed to obtain G, if any, are called identification assumptions for G and are denoted $\mathcal{I}(\Gamma)$.

The notion of inverse problem and the adjacent problem of identification are illustrated in Appendix B where the models in Examples A.1, A.2, A.3, and A.4 under complete observability

are shown to have mirror formulations as inverse problems by appropriately building the right coupling for some well-chosen unobservables. The applicability of the notion is further illustrated in Appendix C where we isolate the inverse problems for the examples in the Introduction with their respective identification assumptions as considered by the authors (with the caveat that the main problem tackled in Example 1.3 is not properly an inverse problem).

To conclude this section, we prove a number of results that complete the description of inverse problems and clarify their relation to forward problems. In particular, we show that:

- any function of the coupling can be targeted in an inverse problem – see Remark 4.2;
- inverse problems need not be connected to inverse problems in operator theory – see Remark 4.3;
- inverse problems induce statistical models in the sense of parametrized families of distributions that lead to the more common definition of identification as in Matzkin (2007) – see Remark 4.4;
- the separation of all constraints on the coupling Γ of an inverse problem (Γ, g) into a preliminary defining set $\mathcal{C}(\Gamma)$ and an identification set $\mathcal{I}(\Gamma)$ is always arbitrary – see Remark 4.5;
- identified inverse problems always induce some forward problems – see Remark 4.6;
- the problem of specification for inverse problems always makes the validity of the inferential results conditional on an unavoidable "leap of faith" which prevents empirical closure – see Remark 4.7;
- sampling is a modeling choice that always constrains the coupling – see Remark 4.8.

Remark 4.2. The object of interest in an inverse problem (Γ, g) can be chosen as any feature of the coupling Γ as made clear in Definition 4.1. This first includes unknown constants in the constraints $\mathcal{C}(\Gamma)$, which often appear in the formulation of a stochastic model. This is the case in all the examples of Appendix B where the objects to be recovered are the unknown constants μ, m, β, and ρ that appear in the respective stochastic models. For the sake of generalization and interpretation, it is best to redefine these inverse problems and directly integrate these unknown constants as constant random variables in the vector U of unobservables – this is what is done in all the examples of Appendix B. The interpretative reason for that redefinition will become clear in Section 4.2 and in Appendix D. It is important to note, however, that unknown constants in $\mathcal{C}(\Gamma)$ are not the only objects that can be targeted in an inverse problem, all functions $g(\Gamma)$ of the coupling Γ can. These include, in particular, the full distribution Γ itself, some of its marginal projections, any of their quantiles or any of their moments (provided they exist). This is illustrated in Examples C.2 and C.3 in Appendix C. The range of possibilities is large but a caveat is in order: it is implicit in Definition 4.1 that the function g is deterministic. Indeed, it is generally unfeasible to take as object of interest an unobservable random variable (defined on the same probability space as those in the model). The only case it is possible to do so is when the random variable is constant because its

distribution is then completely indicative of its only value (leading in particular to a deterministic g). When the unobservable random variable is non-constant, there is no feasible path within the model to get back to the even space Ω needed to characterize the random variable. In fact, taking a (non-constant) random variable as object of interest does not lead to an inferential problem, but to what can be defined as a prediction problem. This is illustrated in Example C.3 in Appendix C.

Remark 4.3. Inverse problems in statistics and econometrics have been historically tied to deterministic inverse problems in the sciences involving the inversion of an operator – see the discussion and the references in the Literature Review. It is first important to distinguish between the two problems – a statistical one and an operator-theoretic one. It is then crucial to note that inverse problems in statistics and econometrics, whose common denominator is the statistical recovery of an indirectly observable object, go beyond the stochastic perturbation of inverse problems in operator theory. Indeed, it is directly seen in Appendix B that, while all these simple examples could be viewed as stochastic perturbations of deterministic systems, not all these deterministic systems form inverse problems from the viewpoint of operator theory – only Example B.4 and Example B.3 do. It is then seen in the more convoluted examples of Appendix C that not all problems defined by the statistical recovery of an indirectly observable object originate from the stochastic perturbation of deterministic systems. The frame provided by Definition 4.1 acknowledges and captures this variety under a single construction, independently of where unobservability stems from, be it measurement error as in Example C.1, unobserved characteristics in consumer tastes as in Example C.2, or the natural of potential outcomes within causal theory as in Example C.3.

Remark 4.4. As forward problems, inverse problems indirectly induce statistical models in the form of parametrized families of distributions. However, these families are rarely, if ever, considered in practice because they are convoluted and play little role beyond the stochastic models used to define them. More importantly, they have a tendency to hide the nature of inverse problems and to mask the role played by unobservables. The rare cases they may appear explicitly are for the simple problems of Appendix B where the objects to be recovered are explicit parameters of the stochastic model and when the problem of identification is handled in an abstract way as in Matzkin (2007). For completeness, we show how it is always possible to obtain these parametrized families from inverse problems. This allows to connect our construction to the one in Matzkin (2007) and prove that the usual notion of identification as a property of the parametrization is equivalent to the condition given in Definition 4.2. For this purpose, let us consider an arbitrary inverse problem (Γ, g) for some distribution P_X on E where the coupling is constrained by a preliminary set of assumptions that we denote $\mathcal{C}(\Gamma)$. This set of constraints may contain both a stochastic model and distributional assumptions. From (Γ, g) and $\mathcal{C}(\Gamma)$, we can define two statistical models: one for the observable marginal P_X and one for the full coupling Γ. If the coupling has values in $E \times F$, then we can define the unindexed families of measures

$$\mathscr{P}_m = \{P \in \mathcal{M}(E) : \exists \Delta \in \mathcal{M}(E \times F) \text{ s.t. } P = \Delta_{|E} \text{ and } \mathcal{C}(\Delta)\}$$

and

$$\mathscr{P}_T = \{\Delta \in \mathcal{M}(E \times F) : \mathcal{C}(\Delta)\}$$

where $\Delta_{|E}(A) = \Delta(A \times F)$ for all $A \in \mathcal{E}$ and $\Delta_{|F}(B) = \Delta(E \times B)$ for all $B \in \mathcal{F}$. The object of interest $\theta_0 = g(\Gamma)$ can then be used as parameter. This leads to two parametrized statistical models

$$\{P^\theta \in \mathscr{P}_m : \theta \in \Theta\}$$

and

$$\{P^\theta \in \mathscr{P}_T : \theta \in \Theta\}$$

where $\Theta = g(\mathscr{P}_T)$. As for forward problems, the parametrization is not necessarily valid for the unindexed sets $\mathscr{P}_m$ and $\mathscr{P}_T$ since the function g may not be bijective. However, they remain valid for some sets $\mathscr{P}'_m$ and $\mathscr{P}'_T$ contained in $\mathscr{P}_m$ and $\mathscr{P}_T$, respectively, that can be obtained by taking equivalent classes under parameter equality[2]. Identification then rewrites in terms of the injectivity of the index map for $\mathscr{P}'_m$ given by

$$\begin{aligned} I\colon \Theta = g(\mathscr{P}_T) &\to \mathscr{P}'_m, \\ \theta = g(\Delta) &\mapsto P = \Delta_{|E} \end{aligned}$$

Indeed, it is not always true that I is injective. To see that, note that if $I(g(\Delta)) = I(g(\Delta'))$ so that $\Delta_{|E} = \Delta'_{|E}$, then it is not always true that $g(\Delta) = g(\Delta')$ since $\Delta_{|F}$ and $\Delta'_{|F}$ may differ when too many degrees of freedom are left in the initial constraints on the coupling Γ from which the inverse problem derives. If the index map I is injective, however, then it is invertible (due to the restriction to $\mathscr{P}'_m$) and we have

$$I^{-1}(\Delta_{|E}) = I^{-1}(I(g(\Delta))) = g(\Delta).$$

This is equivalent to the definition of identification given in Definition 4.2 where the function G is simply taken to be I^{-1}. This condition in terms of injectivity is often taken as the definition of identification. It is, however, rarely linked to practical inverse problems due to the difficulty to properly define statistical models for these problems. The difficulty is handled in the construction above but it also shows, by contrast, the logical primitivity and simplicity of the inverse structure exhibited in Definition 4.1 and of the characterization of identification given in Definition 4.2.

Remark 4.5. An inconvenient fact of inverse problems is that there is no principled difference between assumptions used for preliminarily constraining the coupling Γ when setting up an inverse problem and later assumptions used in the identification step for the recovery of the object of interest. This is illustrated in Examples B.1, B.2, B.3, and B.4. In Examples B.1 and B.2, sufficiently many assumptions on Γ are used in the definition of the inverse problem to make the identification step disappear. In Examples B.3 and B.4, the coupling has not been sufficiently constrained for the direct recovery of $g(\Gamma)$ leading to a necessary identification step in the form of

[2] As far as we know, this technical detail is not considered in Matzkin (2007) and the related literature that builds on it. It is natural, however, to assume it to be of second importance, expect maybe in very pathological cases.

additional assumptions on Γ. For any given inverse problem, it is always possible to redefine it by attaching the constraints on the coupling from the identification step so as to make the identification problem disappear. Conversely, it is always possible to strip constraints from the coupling of an already identified inverse problem to make the object of interest no longer identified. There is always an arbitrariness in inverse problems as to when the identification problem starts.

Remark 4.6. An identified inverse problem always induce a forward problem with structure $(\mathscr{P}, f)$ where f is given by the identification function G and $\mathscr{P}$ is the unindexed family of measures for P_X described in Remark 4.4 when all constraints on the coupling are added, including the identification assumptions. That is,

$$\mathscr{P} = \{P \in \mathcal{M}(E) : \exists \Delta \in \mathcal{M}(E \times F) \text{ s.t. } P = \Delta_{|E} \text{ and } \mathcal{C}(\Delta) \cup \mathcal{I}(\Delta)\}$$

where $\mathcal{C}(\Delta)$ denotes the preliminary constraints on the coupling and $\mathcal{I}(\Delta)$ denotes the set of identification assumptions for G. This forward problem encapsulates all the implied constraints on the observable distribution by the identified inverse problems. The problem of specification for this forward problem captures the idea of "empirical adequacy" of inverse problems as further explained in the next remark.

Remark 4.7. The problem of specification for an inverse problem (Γ, g) relates to the choice of a coupling Γ for the observable phenomenon P_X, which usually corresponds to the choice of: a set of unobservable variables, a stochastic model linking the observables to the unobservables, preliminary distributional assumptions on the coupling, and identification assumptions for the object of interest. Ideally and intuitively, we would want to say that an inverse problem is well-specified if all the constraints on the coupling are empirically valid. This is, however, an unfeasible request because some constraints directly bear on unobservables for which we have, by definition, no observations. This makes the specification of inverse problems liable to an unavoidable "leap of faith": there is necessarily a point where the unobservable structure is not reducible to its observational implications and hence evades empirical support. This "leap of faith" is the direct cost to bear for the theoretical interpretation of the inferred object as made clear in Section 4.2. To make the point more explicitly, we can distinguish two overlapping problems in the specification of inverse problems:

1. a problem of empirical adequacy where one only looks at the implied constraints on the observables – this is equivalent to the problem of specification for the forward problem that is induced by the identified inverse problem as defined in Remark 4.6;

2. a problem of theoretical validity for the model linking observables to unobservables where validity is understood either in a realist sense of correspondence to an underlying reality or in an instrumentalist way as compliance with the rules of a pre-specified game.

To interpret the inferential results as intended, the two problems have to be handled jointly as none is individually sufficient. The unavoidable "leap of faith" in specification comes from the fact that,

for the same observable distribution, any two inverse problems with similar empirical adequacy cannot be distinguished empirically: the choice of one or another has to be done on theoretical grounds alone. One may still hope that empirical adequacy is a good enough guide in most cases (see Spanos and Mayo (2015)). However, it is clear, as we move from the simple examples of Appendix B to the more convoluted ones of Appendix C, that empirical adequacy stops being a guide for the choice of the coupling since almost no assumptions are made on the observables (see Example C.3 in particular). In many bona fide cases, the many possible empirically adequate couplings have to be distinguished from a priori theoretical reasons only. We believe that this point has not been sufficiently emphasized in the literature on the topic given its unbalanced focus on empirical adequacy (see again Van Fraassen (1980) and Spanos and Mayo (2015)). We also believe that this limitation offers good support for a realist interpretation of models: since empirical adequacy is often of no help for model selection, resorting to the most widely shared theoretical language to weigh assumptions made on their unobservable structure is probably the safest choice.

Remark 4.8. As for forward problems, we introduced inverse problems under an i.i.d. framework with dependence and distributional-shift as a special case – see Remark 4.1. This was done for the sake of exposition, but it should not hide the fact that sampling is a modeling choice that implicitly constrains the coupling. This can be made explicit by redefining inverse problems for the distribution $P_{X_1,\dots,X_n}$ of the full observable sample $(X_1, X_2, \dots, X_n)$ instead of the hypothesized identical marginal P_X. The coupling is then implicitly defined for $P_{X_1,\dots,X_n}$ and $P_{U_1,\dots,U_n}$ where $P_{U_1,\dots,U_n}$ is the joint distribution of the full vector $(U_1, U_2, \dots, U_n)$ of hypothesized unobservables. Under this formulation, it is clear that an i.i.d. assumption is the most restrictive constraint on the coupling among all possible sampling choices as it implies the most invariant joint distribution for the observables and the most invariant joint distribution for the unobservables – both some product measure for some identical marginals P_X and P_U, respectively. It is then seen that the sampling choice has important consequences for the explanations sustained by inverse problems in virtue the theory-ladenness made clear in next section. This connection is illustrated with the examples of the Introduction in Appendix D and discussed at more length in Remark D.1.

4.2 Theory-ladenness, leap of faith, and explanation

The previous section made clear that inverse problems are fundamentally different from forward problems – this naturally mirrors the observed and historical variance between the statistical literature and the econometric literature. The fact that forward problems can always be obtained by solving some well-chosen inverse problems, illustrated in all the examples of Appendix B, should not hide this difference. Indeed, the specification of a coupling between distributions that can be sampled from and distributions that cannot commit the modeler to more than a simple set of distributions for the observables. It is the objective of this section to make precise the nature of this commitment. For this purpose, we will answer two intertwined questions:

1. Where do inverse problems originate from?

2. What is gained and what is lost by working with unobservables?

The first question has a simple answer that is distinctly illustrated in all the examples of the Introduction: inverse problems exist because one does not simply want to describe the observations but instead to leverage them to recover or characterize an underlying structure from which the observations are supposed to derive. The hypothesizing of an underlying structure for some observable phenomena is exactly what a "scientific theory" is according to philosophy of science – measurements are made sense of by way of laws and entities that cannot be directly measured. We refer to Van Fraassen (1980) for an in-depth exploration of this question. For our purpose, what matters is that we can legitimately view inverse problems as the statistical version of scientific theories. Their origin can be found either in a deterministic theoretical model that is simply translated in stochastic terms or in directly hypothesizing a probabilistic relations without connection to a preliminary non-random structure. In any case, this exercise always supposes the existence of a field-specific theoretical bedrock that gives a particular meaning and interpretation to all the components of an inverse problem, and in particular to the unobservables. It is this theoretical coating that guides and constrains the selection of a particular coupling. It is the hypothesized probabilistic implications of the theory that are used to assess the plausibility of the assumptions for identification. This is what we refer to as the theory-ladenness of unobservables.

This makes clear what is gained by working with a coupling that cannot be completely sampled from: the possibility to go beyond the observable with the (partial) recovery of an underlying structure, that is, a data-supported explanation. It is the stochastic nature of unobservables in inverse problems construed as scientific theories that allow to directly leverage the statistical apparatus and bring data into theoretical models. The cost of this procedure is the leap of faith characterized in Remark 4.7: the fact that no data is ever available for the unobservable processes prevent explanations to be completely supported by the observations; there is always a range of observationally equivalent models one has to select from based on theoretical reasons only. The problem of identification for inverse problems can then be understood as a formalization of the problem of "underdetermination" of scientific theories from within statistics. This underdetermination, however, is not sufficient to crown empirical adequacy as the sole guide for the selection of scientific theories, at least from the statistical perspective. As made clear in Remark 4.7, many models considered in practice constrain the observables very minimally in favor of assumptions on the unobservables so that empirical adequacy ends up having almost no bite. The success of some of these models, where the properties of unobservables have to be discussed on theoretical grounds alone, may be viewed as a challenge for constructive empiricism at the statistical level. On the other hand, this supposes that statisticians and econometricians pay full attention to the unobservables and their respective theoretical bedrock when exporting the inverse methodology to new problems – this is not always the case in practice where similar assumptions are simply transposed from one theory to another. This begs the question of what are appropriate criteria to judge and compare inverse problems – we leave a full exploration of this question to future research.

To avoid unnecessary conceptual discussions, we close this section by referring to Appendix D where we illustrate all the claims above in the three examples of the Introduction.

5 Conclusion

Model-based inferential statistics always consists in constraining a set of possible distributions for some random variables of which finitely many realizations are assumed accessible. These random variables are naturally called observable and their realizations observations. These observations are then used to characterize some invariant properties of the observable distribution where the invariance stems from the modeling constraints. Our paper makes clear that there exist two different ways to come up with such constraints and that each gives to the respective statistical results very different interpretations. In one case, only the observable random variables are assumed to exist. This naturally limits the range of statistical tasks that can be considered with an upper bound in the recovery of the joint distribution of the observables. The advantage is that, in principle at least, the modeling constraints on the distributions can always be tested against the observations. This task, descriptive and empirically closed, is captured by what we called forward problems. In the other case, additional random variables, for which no realizations are ever accessible but on which the observables depend, are assumed to exist. This embedding invariably ties one's hands by preventing the data to be the sole guide and goal, but it opens the statistical apparatus to a wide range of new questions as it permits the recovery of indirectly observable objects. These problems can be legitimately called inverse because the objective is to "undo" the effects of the unobservables on the object of interest so that it can be recovered from the data. We showed that they can all be handled under a single framework by simply looking at the coupling between observable and unobservable distributions implicitly defined in each case. This formalization simplifies a number of standard issues in model-based inferential problems in the presence of unobservables, including the problem of identification. It also allows us to make clear the dependence of any inverse problems in statistics and econometrics on a preliminary field-specific theory that justifies and constrains it. Unobservables are then seen to be what allows one to bring data into scientific theories through the discipline of statistics to form empirically-supported explanations. The natural cost of this extension of the statistical reach is an unavoidable leap of faith due to the nature of unobservability itself: some assumptions on the coupling can never be empirically supported and require one to argue on theoretical grounds alone. We have remained mostly silent throughout our discussion on what tools would allow one to best weigh, judge, and compare the plausibility of different assumptions on unobservable structures that are severed from empirical support – in practice, we observe an informal back-and-forth between field-specific theoretical justifications and statistical considerations as exemplified, for instance, in the bona fide examples of this paper. Given the theory-dependence of inverse problems, it is unclear if a single framework can be devised to discipline statistical assumptions on the unobservables across fields – it is clear, however, that doing so would significantly help increase trust in the results of models with unobservables.

References

Aster, R. C., B. Borchers, and C. H. Thurber (2018): *Parameter Estimation and Inverse Problems*. Elsevier.

Bahadur, R. R., and L. J. Savage (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *The Annals of Mathematical Statistics*, 27(4), 1115–1122.

Berger, J. O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Springer New York.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Box, G. E. (1976): "Science and Statistics," *Journal of the American Statistical Association*, 71(356), 791–799.

Breiman, L. (2001): "Statistical Modeling: The Two Cultures," *Statistical Science*, 16(3), 199–231.

Cavalier, L. (2011): "Inverse Problems in Statistics," in *Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, August 31-September 4, 2009*, pp. 3–96. Springer.

Cox, D. R. (1990): "Role of Models in Statistical Analysis," *Statistical Science*, 5(2), 169–174.

Evans, S. N., and P. B. Stark (2002): "Inverse Problems as Statistics," *Inverse Problems*, 18(4), R55.

Ferguson, T. S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75(1), 259–276.

Fisher, R. A. (1922): "On the Mathematical Foundations of Theoretical Statistics," *Philosophical transactions of the Royal Society of London. Series A*, 222(594-604), 309–368.

——— (1925): *Statistical Methods for Research Workers*. Oliver & Boyd.

——— (1935): *The Design of Experiments*. Oliver & Boyd.

Florens, J.-P., and A. Simoni (2017): "Introduction to the Special Issue on Inverse Problems in Econometrics," *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 128, 1–3.

Giné, E., and R. Nickl (2021): *Mathematical Foundations of Infinite-dimensional Statistical Models*. Cambridge University Press.

Györfi, L., M. Kohler, A. Krzyzak, and H. Walk (2002): *A Distribution-Free Theory of Nonparametric Regression*. Springer.

Hu, Y., S. Schennach, and J.-L. Shiu (2022): "Identification of Nonparametric Monotonic Regression Models with Continuous Nonclassical Measurement Errors," *Journal of Econometrics*, 226(2), 269–294.

Kennedy, E. H. (2023): "Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects," *Electronic Journal of Statistics*, 17(2), 3008–3049.

Lehmann, E. (1990): "Model Specification: The Views of Fisher and Neyman, and Later Developments," *Statistical Science*, 5(2), 160–168.

Lehmann, E. L., and G. Casella (1998): *Theory of Point Estimation*. Springer.

Lei, L., and E. J. Candès (2021): "Conformal Inference of Counterfactuals and Individual Treatment Effects," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5), 911–938.

Matzkin, R. L. (2007): "Nonparametric Identification," *Handbook of Econometrics*, 6, 5307–5368.

Mayo, D. G. (1997): "Duhem's Problem, the Bayesian Way, and Error Statistics, or "What's Belief Got to Do with It?"," *Philosophy of Science*, 64(2), 222–244.

Moon, H. R., M. Shum, and M. Weidner (2018): "Estimation of random coefficients logit demand models with interactive fixed effects," *Journal of Econometrics*, 206(2), 613–644.

Neyman, J. (1952): *Lectures and Conferences on Mathematical Statistics and Probability*. Graduate School, US Department of Agriculture.

——— (1976): "The Emergence of Mathematical Statistics," in *On the History of Statistics and Probability*, ed. by D. B. Owen. Dekker.

——— (1977): "Frequentist Probability and Frequentist Statistics," *Synthese*, 36(1), 97–131.

Robert, C. P. (2007): *The Bayesian Choice: From Decision-theoretic Foundations to Computational Implementation*. Springer.

Schennach, S. M. (2020): "Mismeasured and Unobserved Variables," in *Handbook of Econometrics*, vol. 7, pp. 487–565. Elsevier.

Shao, J. (2008): *Mathematical Statistics*. Springer.

Spanos, A., and D. G. Mayo (2015): "Error Statistical Modeling and Inference: Where Methodology Meets Ontology," *Synthese*, 192, 3533–3555.

Sudakovand, V., and L. Khalfin (1964): "Statistical approach to ill-posed problems in mathematical geophysics," *Sov. Math.—Dokl*, 157.

Tarantola, A. (2005): *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM.

Van Fraassen, B. C. (1980): *The Scientific Image*. Oxford University Press.

Woodward, J., and L. Ross (2021): "Scientific Explanation," in *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University, Summer 2021 edn.

Appendices

We start with a few notational remarks. All measurable spaces considered in what follows are assumed to be endowed with their Borel σ-algebra. In particular, if E is a Euclidean space, then $\mathcal{M}(E)$ denotes all Borel probability measures on E. If two random variables are introduced without mention of their respective sample space, it is generally assumed that this space is the same and is denoted $(\Omega, \mathcal{B}, \mathbb{P})$. For any measure space $(E, \mathcal{E}, \mu)$, the set $L^p(E, \mathcal{E}, \mu)$ or simply L^p is the Lebesgue space of order p on this measure space. If a measure on an Euclidean space is assumed dominated, the dominating measure is always taken to be the Lebesgue measure.

Appendix A Propaedeutic examples: forward problems

Example A.1 (Non-parametric Mean Estimation)**.** Non-parametric mean estimation consists in recovering the expected value of a real-valued random variable X with distribution P_X based on observations $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ construed as an i.i.d. sample from P_X. This naturally defines a forward problem whose structure follows from the simple consideration that: the descriptive first-best solution would be the complete recovery of P_X; this is a hard statistical problem off which we trade the simpler statical task of recovering its expected value (provided it exists). Non-parametric mean estimation thus forms a forward problem $(\mathscr{P}, f)$ where $\mathscr{P}$ is simply the set of all distributions on $\mathbb{R}$ with finite expected value, that is,

$$\mathscr{P} = \left\{ P \in \mathcal{M}(\mathbb{R}) : \int_{\mathbb{R}} x \, dP(x) < \infty \right\}.$$

and f is the expected value operator, that is,

$$f \colon \mathscr{P} \to \mathbb{R},\ P \mapsto \int_{\mathbb{R}} x \, dP(x).$$

The expected value exists and is unique for each $P \in \mathscr{P}$ so f is a well-defined function and the forward problem is hence properly defined. The forward structure makes clear the trade-off between completeness of the description and difficulty of the statistical problem: potential information is lost by moving from $\mathcal{M}(\mathbb{R})$ to $\mathscr{P}$ and from P to $f(P)$, but the resulting statistical task is made easier. This does not mean, however, that the task is trivial: Bahadur and Savage (1956) showed that, due to the dependence of the mean on the tails, the choice of a large non-parametric set $\mathscr{P}$ prevents the existence of "effective" statistical procedures for $f(P_X) = \mathbb{E}[X]$.

Example A.2 (Mean Estimation in a Gaussian Family)**.** Mean estimation in a Gaussian family with known variance is a sub-problem directly obtained from non-parametric mean estimation covered in Example A.1 by further assuming that P_X is given by $N(\mu_0, 1)$ for some $\mu_0 \in \mathbb{R}$ (where the variance is normalized to 1 without loss of generality). This directly forms a forward problem $(\mathscr{P}_G, f)$ where $\mathscr{P}_G$ a strict subset of $\mathscr{P}$ defined in Example A.1 by restricting it to Gaussian

distributions with variance equal to 1, that is,

$$\mathscr{P}_G = \{N(\mu, 1) : \mu \in \mathbb{R}\},$$

and f is taken as the restriction of the expected value operator to $\mathscr{P}_G$, namely

$$f : \mathscr{P}_G \to \mathbb{R},\ P \mapsto \int_{\mathbb{R}} x\, dP(x).$$

The expected value exists and is unique for every $P \in \mathscr{P}_G$ and so the forward problem is properly defined. The structure again helps make clear the trade-off between descriptive completeness and statistical difficulty: additional information is potentially lost by moving from $\mathscr{P}$ to $\mathscr{P}_G$ but this is exactly what is gained in the simplification of the inferential problem – the Gaussian assumption ensures control of the tails preventing the behaviors in Bahadur and Savage (1956) to manifest. It should be noted that, in this case, the restriction to $\mathscr{P}_G$ implies that there is no informational loss in moving from P to $f(P)$ since the restriction of f to $\mathscr{P}_G$ is now bijective. In other words, all potential information loss in this problem is in restricting $\mathcal{M}(\mathbb{R})$ to $\mathscr{P}_G$. The formalization of Example A.2 being a sub-problem of Example A.1 is made clear in Remark 3.5. The consequences of this ordering for the descriptive scope are illustrated above.

Example A.3 (Non-parametric Regression)**.** Non-parametric regression under complete observability consists in recovering the best L^2-approximation of a square-integrable random variable Y as a function of another square-integrable X random variable based on observations $((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) \in \mathbb{R}^{2n}$ construed as an i.i.d. sample from the joint distribution $P_{X,Y}$ of (X, Y). This directly defines a forward problem whose structure follows from the simple consideration that: the descriptive first-best solution would be the complete recovery of $P_{X,Y}$; we trade this hard statistical problem off for the simpler task of recovering the best L^2-approximation of Y as a function of X (provided it exists). This can be naturally decomposed into a forward problem $(\mathscr{P}, f)$ as follows. The set $\mathscr{P}$ is taken as all distributions on $\mathbb{R}^2$ with finite second moments, that is,

$$\mathscr{P} = \left\{P \in \mathcal{M}(\mathbb{R}^2) : \int x^2\, dP_1(x) < \infty,\ \int x^2\, dP_2(x) < \infty\right\},$$

where $P_1(A) = P(A \times \mathbb{R})$ and $P_2(A) = (\mathbb{R} \times A)$ for all $A \in \mathcal{B}(\mathbb{R})$. The function f is then taken as the projection given by

$$f : \mathscr{P} \to \{m \in \{\mathbb{R} \to \mathbb{R}\} : m \text{ is measurable}\},$$
$$P \mapsto \arg\min_{m \text{ meas.}} \int (y - m(x))^2\, dP(x, y).$$

A solution to the optimization problem is known to exist and corresponds to the conditional expectation of Z_2 given Z_1 where $Z_1 \sim P_1$ and $Z_2 \sim P_2$. In particular, we have $f(P_{X,Y}) = m^*(\cdot) = \mathbb{E}[Y|X = \cdot]$ using standard notations. However, a technical caveat is in order: the conditional

expectation is not unique on negligible sets. This implies that f is not a function as defined above, which prevents $(\mathscr{P}, f)$ to be a proper forward problem. To solve this technical issue, one standard solution is to modify the co-domain of f from measurable functions to equivalent classes of functions that only differ on non-negligible sets. With this modification, f is a proper function and the forward problem $(\mathscr{P}, f)$ for non-parametric regression is properly defined. This decomposition appropriately captures the trade-off between descriptive completeness and statistical difficulty: a potential informational loss occurs by moving from $\mathcal{M}(\mathbb{R}^2)$ to $\mathscr{P}$ and from P to $f(P)$, but the statistical problem is made simpler as a result. The task is, however, far from trivial as exemplified, for instance, in Györfi, Kohler, Krzyzak, and Walk (2002).

Remark A.1. A weaker version of non-parametric regression can be considered by simply assuming finite absolute first moments of the marginals in the definition of $\mathscr{P}$ and directly taking the conditional expectation operator defined on L^1 in the definition of f. This generalization comes at the cost of the geometric interpretation of the problem.

Example A.4 (Linear Regression)**.** Linear regression belongs to the same class of problems as non-parametric regression in the sense that they consider the same set of possible distributions. To see this, we decompose it as a forward problem $(\mathscr{P}, f)$ where $\mathscr{P}$ is again given by

$$\mathscr{P} = \left\{P \in \mathcal{M}(\mathbb{R}^2) : \int x^2 \, dP_1(x) < \infty, \int x^2 \, dP_2(x) < \infty\right\}.$$

The difference with non-parametric regression stems from the choice of f. The idea is to restrict attention to the best linear L^2-approximation of Y as a function of X among all such L^2-approximations. That is, f is now chosen as

$$\begin{aligned} f\colon \mathscr{P} &\to \left\{m \in \{\mathbb{R} \to \mathbb{R}\} : m \text{ is linear}\right\}, \\ P &\mapsto \arg\min_{m \text{ linear}} \int (y - m(x))^2 \, dP(x, y). \end{aligned}$$

By extending the definition of f to equivalence classes of functions as in non-parametric regression, the map f becomes a well-defined function and the modified forward problem is hence properly defined. This forward structure captures again the trade-off between descriptive completeness and statistical simplicity: by restricting f to linear solutions, more information is potentially lost; this leads, however, to a more manageable statistical problem. To see this, it is useful to re-frame the problem as commonly done in the literature on linear regression. Indeed, the projection problem given by f in linear regression is directly seen to be equivalent to

$$\bar{f}\colon \mathscr{P} \to \mathbb{R}, \quad P \mapsto \arg\min_{\beta \in \mathbb{R}} \int (y - x\beta)^2 \, dP(x, y).$$

This simplification makes clear the dimensionality reduction in comparison to non-parametric regression. There is, however, a subtlety in this re-framing since the function $\bar{f}$ may not be well-defined anymore due to the fact that the first-order condition $\int (y - x\beta)x \, dP(x, y) = 0$ may have

several solutions. If we want to keep the co-domain of $\bar{f}$ as $\mathbb{R}$, then the set of possible descriptions can be changed to

$$\bar{\mathscr{P}} = \left\{P \in \mathscr{P} : \int_{\mathbb{R}} x^2 \, dP_1(x) \neq 0\right\}.$$

Indeed, the condition $\int x^2 \, dP_1(x) \neq 0$ then ensures that the minimizer in the definition of $\bar{f}$ is unique with closed-form expression $\beta^* = (\int x^2 \, dP_1(x))^{-1} \int xy \, dP(x, y)$. This makes $\bar{f}$ a proper function and $(\bar{\mathscr{P}}, \bar{f})$ a proper forward problem. The idea that linear regression and non-parametric regression are two approximations, one weaker than the other, to the same first-best description given by $P_{X,Y}$ is formalized in the definition of classes introduced in Remark 3.5.

Remark A.2. Another solution to the identification issue in the re-framed linear regression problem is to change the co-domain of $\bar{f}$ from $\mathbb{R}$ to sets of $\mathbb{R}$. The solutions would be singletons whenever $\int x^2 \, dP_1(x) \neq 0$ and the whole set otherwise. This alternative is a valid option but rarely considered in practice for linear regression. It is, however, important to note that it is the actual counterpart to taking equivalent classes of functions in the original formulation. To draw attention to the difference, identification where G has values in sets are often referred to as set identification. This still falls under the general frame of Definitions 4.1 and 4.2 whatever the denomination.

Example A.5 (Auto-Correlation Estimation)**.** Auto-correlation estimation is a standard problem in the analysis of time series. We consider in this example the recovery of the first-order auto-correlation of a vector $X = (X_1, X_2, \ldots, X_T)$ taking values in $\mathbb{R}^T$ that is assumed mean-zero stationary in the sense defined below. The available data $(x_1, x_2, \ldots, x_T)$ consists of a sample of size 1 drawn from the distribution P_X of X. This naturally defines a forward problem whose forward structure derives from the simple consideration that: the descriptive first-best would be to recover the distribution P_X of the observable random vector X; however, the problem is not tractable from the available data and we tackle instead the simpler task of recovering the first-order correlation structure of the series by assuming it is invariant. This forces us first to restrict the possible descriptions for X by taking $\mathscr{P}$ as the set of distributions $P \in \mathcal{M}(\mathbb{R}^T)$ such that:

1. $\int x \, dP_t(x) = 0$ for all $t \in \{1, 2, \ldots, T\}$;
2. there exists $\sigma^2 > 0$ such that $\int x^2 \, dP_t(x) = \sigma^2$ for all $t \in \{1, 2, \ldots, T\}$;
3. there exists $\gamma \in \mathbb{R}$ such that $\int xy \, dP_{s,s+1}(x, y) = \gamma$ for all $s \in \{1, 2, \ldots, T-1\}$;

where P_t denotes the marginal of the t^{th} entry of P and $P_{s,s+1}$ the joint of the s^{th} and $(s+1)^{\text{th}}$ entries of P. This stationary structure allows us to meaningfully recover the first-order auto-correlation coefficient γ/σ^2 for the series. Given the restrictions on the descriptions, this defines a valid forward problem $(\mathscr{P}, f)$ where f maps any $P \in \mathscr{P}$ to its first-order auto-correlation coefficient γ/σ^2, that is,

$$f \colon \mathscr{P} \to \mathbb{R}, \quad P \mapsto \gamma/\sigma^2.$$

This decomposition in a time series context helps make clear the trade-off between descriptive completeness and statistical difficulty: information is possibly lost by moving from $\mathcal{M}(R^T)$ to $\mathscr{P}$ and from P to $f(P) = \gamma/\sigma^2$, but the statistical problem is made feasible as a result.

Remark A.3. We have not discussed sampling directly as a restriction on the possible descriptions in all the previous examples. It is, however, a fundamental one as made clear in Remark 3.6. In particular, the i.i.d. hypothesis made in Examples A.1, A.2, A.4, and A.3, is a strong constraint on the possible descriptions: it implies the most invariant joint distribution for the full sample. This is seen most strikingly when comparing these examples to Example A.5: independence among the observables under an i.i.d. hypothesis would make the problem of Example A.5 trivial with auto-correlation of order 1 always equal to 0. To make explicit the strong invariance implied by any i.i.d. assumption, it is always possible to re-frame forward problems for the joint distribution of the full sample as made clear in Remark 3.6. This reformulation should be kept in mind whenever one weighs different modeling choices for forward problems in practice.

Appendix B Propaedeutic examples: inverse problems

Example B.1 (White Noise Model)**.** Non-parametric mean estimation under complete observability as considered in Example A.1 has a mirror formulation as an inverse problem with stochastic model known as the white noise model. The stochastic model is given by

$$\mathrm{SM}(X, (\mu, \varepsilon)) \colon \; X = \mu + \varepsilon,$$

where X is some observable random variable taking values in $\mathbb{R}$ and $U = (\mu, \varepsilon)$ is an unobservable random vector with values in $\mathbb{R}^2$. It is further assumed that μ is constant and ε satisfies $\mathbb{E}[\varepsilon] = 0$. The only available data for the problem $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is construed as an i.i.d. sample drawn from P_X. The stochastic model and the assumptions on U constrain a coupling Γ for P_X and P_U from which the inverse problem derives. The object of interest is taken as $g(\Gamma) = \mu$. From the assumptions on the coupling Γ implied by the stochastic model and the assumptions on U, we directly have $\mu = \mathbb{E}[X]$, and so the inverse problem is identified without any additional assumption. In this case, the identifying function G is simply the mean operator applied to P_X. The statistical task is then the same as in Example A.1: it consists in recovering $\mathbb{E}[X] = \mu$ from the observations $(x_1, x_2, \ldots, x_n)$ construed as an i.i.d. sample from P_X.

Example B.2 (Gaussian White Noise Model)**.** Normal mean estimation under complete observability as considered in Example A.2 has also a mirror formulation as an inverse problem with stochastic model known as the Gaussian white noise model. This problem is very similar to the white noise model of Example B.1: the stochastic model has the same additive structure but ε is now assumed to be normally distributed with mean zero and known variance (normalized to 1). That is,

$$\mathrm{SM}(X, (\mu, \varepsilon)) \colon \; X = \mu + \varepsilon,$$

where now it is assumed that $\varepsilon \sim N(0, 1)$ while μ is still taken constant. This model constrains a coupling Γ for P_X and P_U. In this case, the restrictions on Γ have some direct parametric

implications on P_X, namely

$$P_X = N(\mu, 1).$$

As in the white noise model, the object of interest is taken as $g(\Gamma) = \mu$. For the same reasons, it is directly seen that $\mu = \mathbb{E}[X]$ and so the inverse problem is identified with G again the mean operator applied to P_X. The statistical task is then the same as in Example A.2: it consists in recovering μ from the observations $(x_1, x_2, \ldots, x_n)$ construed as an i.i.d. sample from $N(\mu, 1)$.

Example B.3 (Non-parametric Regression Model)**.** Non-parametric regression under complete observability as considered in Example A.3 has a mirror formulation as an inverse problem. The stochastic model defining non-parametric regression as an inverse problem is given by

$$\mathrm{SM}((X, Y), (m, \varepsilon)) \colon\ Y = m(X) + \varepsilon,$$

where (X, Y) is a vector of observable random vectors taking values in $\mathbb{R}^2$ and $U = (m, \varepsilon)$ is a vector of unobservable random variables with values in $M \times \mathbb{R}$ for M some functional space in $\{\mathbb{R} \to \mathbb{R}\}$. It is further assumed that m_0 is constant. The only available data for the problem $((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)) \in \mathbb{R}^{2n}$ is construed as an i.i.d. sample from the joint distribution $P_{X,Y}$ of (X, Y). The stochastic model constrains a coupling Γ for $P_{X,Y}$ and P_U from which the inverse problem derives. The object of interest is taken as $g(\Gamma) = m$ and corresponds to the constant functional relation between Y and X. The problem is, however, not identified as such. A standard set of identification assumptions consists in hypothesizing that:

1. $X, Y, \varepsilon \in L^2(\Omega, \mathcal{B}, \mathbb{P})$;
2. $\mathbb{E}[\varepsilon | X] = 0$.

These assumptions are enough to identify $m(\cdot)$ where G obtains as the conditional expectation operator. Indeed, it is directly seen from the law of iterated expectations that $m(\cdot) = \mathbb{E}[Y | X = \cdot]$. From a technical viewpoint, however, the function $m(\cdot)$ is not exactly identified because the operator G when defined on M has unique images only almost surely. To obtain proper identification, one solution is to modify slightly the stochastic model by taking for $m(\cdot)$ not a function but an equivalent class of functions that are equal on non-negligible sets. The statistical task is then the same as in Example A.3 and consists in recovering $\mathbb{E}[Y | X = \cdot] = m(\cdot)$ from the observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ construed as an i.i.d. sample from $P_{X,Y}$.

Remark B.1. The L^2 condition in the previous example is used to match Example A.3. However, a more natural set of identification conditions in the inverse formulation of non-parametric regression consists simply in assuming that ε and Y have finite expectations and that X, Y, and ε are defined on the same probability space. Indeed, the geometric interpretation of the problem is secondary given the stochastic model, while the relaxation makes the identification assumption more plausible.

Example B.4 (Linear Regression Model)**.** Linear regression under complete observability as considered in Example A.4 has a dual formulation as an inverse problem that bears close resemblance

to the non-parametric regression model of Example B.3. In this case, the functional relation between Y and X is assumed to be linear. This leads to the stochastic model given by

$$\mathrm{SM}((X,Y),(\beta,\varepsilon)\colon\ Y=\beta X+\varepsilon\,,$$

where $U=(\beta,\varepsilon)$ is some unobservable random vector with values in $\mathbb{R}^2$ and β is assumed constant. This constrains a coupling Γ for $P_{X,Y}$ and P_U. As in Example A.4, the object of interest is directly taken to be $g(\Gamma)=\beta$, and not the measurable functions $x\mapsto\beta x$ as in the non-parametric case. This inverse problem (Γ,g) is not identified without additional assumptions. A standard set of identification assumptions consists in further constraining the coupling Γ by hypothesizing that:

1. $X,Y,\varepsilon\in L^2(\Omega,\mathcal{B},\mathbb{P})$;
2. $\mathbb{E}\left[X\varepsilon\right]=0$;
3. $\mathbb{E}\left[X^2\right]\neq 0$.

It is then easily seen that β is identified by taking $G\colon P_{X,Y}\mapsto(\mathbb{E}\left[X^2\right])^{-1}\mathbb{E}\left[YX\right]$. The statistical problem is again the same as in Example A.4 and consists in recovering $(\mathbb{E}\left[X^2\right])^{-1}\mathbb{E}\left[YX\right]$ from the observations $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ construed as an i.i.d. sample from $P_{X,Y}$.

Example B.5 (Auto-Regressive Model)**.** Auto-correlation estimation under complete observability as considered in Example A.5 has a mirror formulation as an inverse problem. The stochastic model for this problem is given by

$$\mathrm{SM}(X,(\rho,\varepsilon))\colon\ X_t=\rho X_{t-1}+\varepsilon_t\quad\text{ for all }\quad t\in\mathbb{Z}\,,$$

where ρ is a random variable in $\mathbb{R}$ which is further assumed constant and $X=(X_t)_{t\in\mathbb{Z}}$ and $\varepsilon=(\varepsilon_t)_{t\in\mathbb{Z}}$ are stochastic processes whose component each takes values in $\mathbb{R}$. The random variables ε and ρ are taken as unobservable. The stochastic process X is assumed observable in the sense that a finite connected set of its components is assumed observable, say, $X^T=(X_1,X_2,\ldots,X_T)$ for some $T\in\mathbb{N}$. The available data is then a sample of size 1 drawn from P_{X^T}. The stochastic model thus constrains a coupling Γ for P_{X^T} and P_{U^T} where P_{U^T} is the joint distribution of $U^T=(\rho,(\varepsilon_2,\varepsilon_3,\ldots,\varepsilon_T))$. This defines an inverse problem (Γ,g) where the object of interest is taken as $g(\Gamma)=\rho$. A standard set of identification assumptions for ρ is given by constraining the initial stochastic process ε and the constant ρ as follows:

1. $\mathbb{E}\left[\varepsilon_t\right]=0$ for all $t\in\mathbb{Z}$;
2. $\mathbb{E}\left[\varepsilon_t^2\right]=\sigma^2<\infty$ for all $t\in\mathbb{Z}$;
3. $\mathbb{E}\left[\varepsilon_t\varepsilon_s\right]=0$ for all $t,s\in\mathbb{Z}$ such that $t\neq s$;
4. $|\rho|<1$.

The condition $|\rho| < 1$ ensures stationarity of the stochastic process X as can be seen by recursion until an infinite moving average representation for X_t is obtained. It is then easily seen that ρ is identified and is equal to the auto-correlation coefficient (of order 1) for X. The statistical problem is then the same as in Example A.5 and consists in recovering $\rho = \gamma/\sigma^2$ from a sample $(x_1, x_2, \ldots, x_T)$ of size 1 drawn from P_{X^T} where P_X satisfies $\mathbb{E}[X_t] = 0$ for all $t \in \mathbb{Z}$, $\mathbb{E}[X_t X_{t+1}] = \gamma$ for some $\gamma \in \mathbb{R}$ for all $t \in \mathbb{Z}$, and $\mathbb{E}[X_t^2] = \sigma^2$ for some $\sigma^2 > 0$ for all $t \in \mathbb{Z}$.

Remark B.2. We framed the previous problem in terms of infinite sequences $X = (X_t)_{t\in\mathbb{Z}}$ and $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ as it is the standard way to introduce it in the literature. However, it is also possible to frame it directly for two finite processes $(Y_1, \ldots, Y_T)$ and $(\varepsilon_2, \ldots, \varepsilon_T)$ so as to exactly match the forward structure of Example A.5. In this case, the condition $|\rho| < 1$ may be changed directly for an hypothesis that the series is stationary in the sense defined in Example A.5.

Appendix C Bona fide examples: inverse problems

Example C.1 (Hu, Schennach, and Shiu (2022))**.** We make explicit the inverse problem in Hu, Schennach, and Shiu (2022) as well as the identification assumptions considered by the authors. By hypothesizing the model

$$Y = m_0(X^*) + \eta\,,$$
$$X = X^* + \varepsilon,$$

the authors implicitly constrain a coupling Γ between some observable random vector (X, Y) with values in $\mathbb{R}^2$ and some unobservable random vector $U = (m_0, X^*, \eta, \varepsilon)$ with values in $M \times \mathbb{R}^3$ for some functional space M in $\{\mathbb{R} \to \mathbb{R}\}$. It is further assumed that m_0 is constant. The object of interest is taken by the authors as $g(\Gamma) = (m_0, f_\varepsilon)$ where f_ε denotes the density of ε (provided it exists). This defines an inverse problem (Γ, g) for which the authors consider the following identification assumptions:

1. The distribution of (X, X^*) admits a density f_{X,X^*} with respect to the Lebesgue measure;
2. The conditional density $f_{X|X^*}$ and the marginal density f_{X^*} are bounded by some constant;
3. η is independent of X^*;
4. $\mathbb{E}[\eta|X^*] = 0$;
5. $\mathbb{E}[\exp(i\gamma\eta)] \neq 0$ for all $\gamma \in \mathbb{R}$;
6. X is independent of Y conditional on X^*;
7. For any function $h \in L^2(\mathcal{X}^*)$, $\int f_{X|X^*}(x|x^*)h(x^*)\,dx^* = 0$ for all $x \in \mathcal{X}$ implies $h(x^*) = 0$ for almost any $x^* \in \mathcal{X}^*$;

8. For any function $h \in L^2(\mathcal{X})$, $\int f_{X|X^*}(x|x^*)h(x)\,dx = 0$ for all $x^* \in \mathcal{X}^*$ implies $h(x) = 0$ for almost any $x \in \mathcal{X}$;

9. There exists a known functional H such that $H[f_{X|X^*}(|x^*)] = x^*$ for any $x^* \in \mathcal{X}^*$;

10. m_0 is continuous, bounded, and strictly monotonic over $\mathcal{X}*$.

The sets $\mathcal{X}, \mathcal{X}^*$ and $\mathcal{Y}$ denote the support of the distributions of X, X^*, and Y, respectively. The main result of the paper (Theorem 2.1) shows that these conditions are sufficient to identify $g(\Gamma) = (m_0, f_\varepsilon)$. In particular, the authors implicitly define a function G as solution to an integral equation that uniquely characterizes $g(\Gamma)$ from $P_{X,Y}$. For the sake of illustration, it should be noted that Assumption (4) leads to $m_0(\cdot) = \mathbb{E}[Y|X^* = \cdot]$ as in Example B.3. However, this expression is not sufficient to identify m_0 here because X^* is unobservable in this problem. Isolating (Γ, g) and the identification constraints on Γ helps make clear the underlying structure that precedes the constrained observable distribution $P_{X,Y}$. It shows, in particular, that recovering $(m_0, f_\varepsilon) = G(P_{X,Y})$ does not only capture information on $P_{X,Y}$ but characterizes the whole system Γ in what forms an empirically-supported explanation. The price to pay for this is an unavoidable leap of faith that stems from unobservability itself as explicit from all the assumptions put on the unobservable structure. The exact nature of the explanation for the identified inverse problem depends on a preliminary field-dependent theory that gives a certain meaning to Γ, to the unobservables, and to the object of interest (m_0, f_ε). This theoretical bedrock is also used when weighing the assumptions on the unobservables severed from empirical support. This is made clear in Appendix D.1 following the general discussion of Section 4.2.

Example C.2 (Moon, Shum, and Weidner (2018))**.** We make explicit the inverse problem in Moon, Shum, and Weidner (2018) as well as the identification assumptions considered by the authors. By hypothesizing the model

$$\delta_{jt} = X'_{jt}\beta^0 + \lambda_j^{0\prime} f_t^0 + e_{jt},$$

$$S_{jt} = \int \pi(\delta_t, X_t, v)\, dG(v; \alpha^0),$$

for all $j = 1, 2, \ldots, J$ and all $t = 1, 2, \ldots, T$ where

$$\pi(\delta_t, X_t, v) = \frac{\exp(\delta_{jt} + X'_{jt}v)}{1 + \sum_{l=1}^{J} \exp(\delta_{lt} + X'_{lt}v)},$$

and G is a known distribution up to α^0, the authors implicitly constrain a coupling Γ between some observable random vector (X, S, Z) with values in $\mathbb{R}^{K\times J\times T} \times \mathbb{R}^{J\times T} \times \mathbb{R}^{M\times J\times T}$ and some unobservable random vector $U = (\alpha^0, \beta^0, \lambda^0, f^0, e, \delta)$ with values in $\mathbb{R}^L \times \mathbb{R}^K \times \mathbb{R}^{J\times R} \times \mathbb{R}^{T\times R} \times \mathbb{R}^{J\times T} \times \mathbb{R}^{J\times T}$. It is assumed that the components of $\alpha^0, \beta^0, \lambda^0$, and f^0 are constant. The object of interest is taken by the authors to be $g(\Gamma) = (\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$. This defines an inverse problem (Γ, g) for which the authors consider the following identification assumptions:

1. For each $t = 1, 2, \ldots, T$ and any value $\alpha \in \mathcal{B}_\alpha \subseteq \mathbb{R}^L$, the equation $S_{jt} =$

$\int \pi(\delta_t, X_t, v)\, dG(v;\alpha)$ can be inverted so that δ_t can be expressed as a unique function of α, X_t, and S_t, which we denote $\delta_t(\alpha)$;

2. $\delta_{jt}(\alpha)$, X_{jt}, and Z_{jt} have finite second moments for all $t = 1, 2, \ldots, T$ and all $j = 1, 2, \ldots, J$;

3. $\mathbb{E}[e_{jt}] = 0$ for all $t = 1, 2, \ldots, T$ and all $j = 1, 2, \ldots, J$;

4. $\mathbb{E}[X_{jt} e_{jt}] = 0$ for all $t = 1, 2, \ldots, T$ and all $j = 1, 2, \ldots, J$;

5. $\mathbb{E}[Z_{jt} e_{jt}] = 0$ for all $t = 1, 2, \ldots, T$ and all $j = 1, 2, \ldots, J$;

6. $\mathbb{E}[(x, z)'(\mathbb{1}_T \otimes M_{(\lambda,\lambda^0)}(x, z)] \geq b\mathbb{1}_{K+M}$ for some $b > 0$ and all $\lambda \in \mathbb{R}^{J\times R}$;

7. $\mathbb{E}[\Delta\xi'_{\alpha,\beta}(x, z)]\mathbb{E}[(x, z)'(x, z)]^{-1}\mathbb{E}[(x, z)'\Delta\xi_{\alpha,\beta}] > \mathbb{E}[\Delta\xi'_{\alpha,\beta}(\mathbb{1}_T \otimes P_{(\lambda,\lambda^0)})\Delta\xi_{\alpha,\beta}]$ for all $(\alpha, \beta) \neq (\alpha^0, \beta^0)$ and all $\lambda \in \mathbb{R}^{J\times R}$;

where vec denotes vectorization of a matrix by stacking its columns, x is a $JT \times K$ matrix with columns $x_{\cdot,k} = \text{vec}(X_k)$, z is a $JT \times M$ matrix with columns $z_{\cdot,m} = \text{vec}(Z_m)$, $\delta^{\text{vec}}(\alpha) = \text{vec}(\delta(\alpha))$, $\Delta\xi_{\alpha,\beta} = \delta^{\text{vec}}(\alpha) - \delta^{\text{vec}}(\alpha^0) - x(\beta - \beta^0)$, $P_{(\lambda,\lambda^0)}$ and $M_{(\lambda,\lambda^0)}$ are the $J \times J$ matrices that project onto and orthogonal to the span of (λ, λ^0). The authors then prove (Theorem 3.1.) that these conditions are sufficient to identify $g(\Gamma) = (\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$. In particular, the authors implicitly define a function G as solution to a well-chosen optimization problem that uniquely characterizes $g(\Gamma)$ from $P_{X,S,Z}$. Isolating (Γ, g) and the identification constraints on Γ helps make clear the underlying structure that precedes the constrained observable distribution $P_{X,S,Z}$. It shows, in particular, that recovering $(\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z}) = G(P_{X,S,Z})$ does not only capture information on the observable distribution $P_{X,S,Z}$ but characterizes the whole system Γ in what forms, in fine, an empirically-supported explanation. The price to pay for this is an unavoidable leap of faith as explicit from all the assumptions put on the unobservable structure that are severed from empirical support. The exact nature of the explanation for the identified inverse problem depends on a preliminary field-dependent theory that gives a certain meaning to Γ, to the unobservables, and to the object of interest $(\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$. This theoretical basis is also used when weighing the identification assumptions on the coupling. This is made clear for the model above in Appendix D.2 following the general discussion of Section 4.2.

Remark C.1. In Example C.2, the observable distribution considered for the problem is the whole distribution of the full sample. This is typical for problems with dependence and distributional-shift as in Example B.5. In this case, however, there is no direct assumption of stationarity or distributional regularity as compared to Example B.5 – in particular, the unobservables λ_j^0 and f_t^0 possibly differ for each $j = 1, 2, \ldots, J$ and each $t = 1, 2, \ldots, T$. This creates a conceptual gap where identification is severed from estimation. This conceptual incongruity has been recognized in the literature but, as far as we know, has never been properly addressed.

Example C.3 (Lei and Candès (2021))**.** We comment on the problem considered in Lei and Candès (2021) and compare it to common inverse problems in causal inference. By considering

the stochastic model

$$Y = Y(1)T + Y(0)(1 - T),$$

the authors implicitly define a coupling Γ between the distribution of the observable random vector (Y, X, T) taking values in $\mathbb{R} \times \mathbb{R}^d \times \{0, 1\}$ and the distribution of the unobservable random vector $U = (Y(1), Y(0))$ taking values in $\mathbb{R}^2$. The object of interest chosen by the authors is the random variable $\tau = Y(1) - Y(0)$ which is not assumed to be constant. This choice, however, does not lead to an inverse problem as made clear in Remark 4.2: it is implicit in Definition 4.2 that g is a deterministic function; the reason is that taking a non-constant random variable as object of interest prevents, de facto, identification in the traditional sense since there is no possibility, in general, to access the sample space Ω on which the random variable is defined. In fact, taking a non-constant random variable as the object of interest constitutes what can be understood as a prediction problem (for which we do not intend to provide a precise definition here), not an inferential problem. It is actually the objective of the authors to move beyond inferential problems for causal inference – they contrast their approach with the more traditional inverse problems considered in the literature. For completeness, we introduce one of these problems, the most elementary one, and briefly comment on two other ones that are considered in Lei and Candès (2021). The main idea is that, while a non-constant $\tau = Y(1) - Y(0)$ cannot be inferred but only predicted, it is still possible to consider as object of inference its distribution or its moments. This would then form a valid inverse problem for Γ provided the moments exist (which we assume in what follows). This is what is done in the most elementary inverse problem in causal inference: the object of interest is taken as the first moment of τ, that is,

$$\tau_{\text{ATE}} = \mathbb{E}[\tau] = \mathbb{E}[Y(1) - Y(0)],$$

which is referred to as the average treatment effect (ATE). Taking $g(\Gamma) = \tau_{\text{ATE}}$ then defines a proper inverse problem (Γ, g) for which identification may be meaningfully considered. A standard set of identifications consists in assuming that:

1. The variables $Y, Y(0), Y(1)$ have finite expectations;
2. $\mathbb{E}[Y(1)|T = 1, X] = \mathbb{E}[Y(1)|X]$;
3. $\mathbb{E}[Y(0)|T = 0, X] = \mathbb{E}[Y(0)|X]$.

This suffices for the identification of τ_{ATE} as it directly follows from mean independence and the law of iterated expectations that

$$\tau_{\text{ATE}} = \mathbb{E}[\mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X]],$$

implicitly defining a function G from $P_{Y,X,T}$ to $\mathbb{R}$ where $G(P_{Y,X,T}) = \tau_{\text{ATE}}$. This inverse problem is the most standard one in causal inference, but it is not only one for the coupling Γ obtained from the binary Neyman–Rubin model. Two standard inverse problems, which are considered in Lei and Candès (2021), consist in the recovery of the conditional average treatment effect (CATE) and

of the conditional quantile treatment effect (CQFT) respectively defined as the functions

$$m : x \mapsto \mathbb{E}\left[Y(1) - Y(0) | X = x\right]$$

and

$$q : x \mapsto Q_\alpha(Y(1)|X = x) - Q_\alpha(Y(0)|X = x)$$

where $Q_\alpha(\cdot|\cdot)$ denotes the α^{th} conditional quantile. We refer to Kennedy (2023) and Firpo (2007), respectively, for identification assumptions and additional statistical considerations for these problems. The isolation of the common coupling Γ in these problems helps make clear the existence of an underlying structure that precedes the recovery of τ_{ATE}, $m(\cdot)$, or $q(\cdot)$ from the observable distribution $P_{Y,T,X}$. In the case of the average treatment effect, for instance, it is clear that recovering $G(P_{Y,X,T}) = \tau_{\text{ATE}}$ does not provide information only about $P_{Y,X,T}$ but about the whole hypothesized system Γ in what forms in fine an empirically-supported causal claim. The price to pay for this causal capability is an unavoidable "leap of faith" that comes from unobservability itself, here the incapacity to completely observe the potential outcomes (which is referred to as the "fundamental problem of causal inference" in this context). This "leap of faith" in the case of the average treatment effect consists mainly in the untestable mean independence conditions (2) and (3) – the severance of this condition from empirical support will be made clearer in Appendix D.3 where we will also discuss the preliminary causal theory that gives to the stochastic model, the potential outcomes, and the average treatment effect their causal interpretation. This dependence on a preliminary field-dependent theory, here a causal theory, is not specific to this problem but common to all inverse problems as made clear in Section 4.2.

Remark C.2. The identification assumptions (2) and (3) for the average treatment effect are often changed for a stronger assumption called "strong ignorability" which consists in assuming that the random vector $(Y(1), Y(0))$ is independent of T conditional on X. In this case, however, one needs to also assume a "no overlap" condition given by $\mathbb{P}(0 < \mathbb{P}(T = 1|X) < 1) = 1$. This is needed so that conditioning on $T = 1$ and $T = 0$ is properly defined. This was not necessary in the previous set of conditions because finiteness of the expectation of $Y(0)$ and $Y(1)$ implies directly from the equalities in (2) and (3) that the conditional expectations on $T = 1$ and $T = 0$ are properly defined as regular conditional probabilities. The inclusion or exclusion of the "no overlap" condition is often a source of confusion in the literature.

Remark C.3. The Neyman–Rubin model where the potential outcomes are unobservable random variables fits within what is often referred to as the super-population approach to causal inference. For completeness, it is important to note that there exists an alternative approach to causal inference – known as the design-based approach – that leverages the Neyman–Rubin model but where the potential outcomes are taken as fixed unobservable values for each observation of the sample. This approach leads to a stochastic model with an extreme form of distributional shift akin to the model

of Example C.2. The stochastic model reads as

$$Y_i = y_i(1)T + y_i(0)(1-T)$$

for all $i = 1, 2, \ldots, n$. The unobservables are taken as $U = (\mathbf{y}(\mathbf{1}), \mathbf{y}(\mathbf{0}))$ where $\mathbf{y}(\mathbf{1}) = (y_1(1), y_2(1), \ldots, y_n(1))$ and $\mathbf{y}(\mathbf{0}) = (y_1(0), y_2(0), \ldots, y_n(0))$ each with values in $\mathbb{R}^n$. It is further assumed that each component of $\mathbf{y}(\mathbf{1})$ and $\mathbf{y}(\mathbf{0})$ is constant. This constrains a coupling Γ for the observable distribution $P_{(Y_1,X_1,T_1),\ldots,(Y_n,X_n,T_n)}$ of the full sample $(Y_1, X_1, T_1), (Y_2, X_2, T_2) \ldots, (Y_n, X_n, T_n)$ and the (degenerate) unobservable distribution P_U. Due to the extreme form of distributional shift, it is generally impossible to take $g(\Gamma) = (\mathbf{y}(\mathbf{1}), \mathbf{y}(\mathbf{0}))$ as the object of interest and obtain an inverse problem that can be identified under reasonable assumptions. With enough constraints on Γ, however, it is still possible to consider a number of valid statistical problems, in particular when it is assumed that the distribution of T is known. In this case, for instance, it is possible to meaningfully test the Fisher null hypothesis that $\mathbf{y}(\mathbf{1}) - \mathbf{y}(\mathbf{0}) = (0, \ldots, 0)$. The structure of Definition 4.1 is seen here to be helpful as a disciplining device to bridge the sampling-based and the design-based approaches to causal inference. The complete exploration of this bridge through inverse problems is left for future research (as is the similarity of design-based causal inference with other models presenting extreme forms of distributional shift such as the factor model of Example C.2).

Appendix D Inverse problems as scientific theories: examples

Example D.1 (Hu, Schennach, and Shiu (2022))**.** We made explicit in Appendix C.1 the inverse problem behind the model in Hu, Schennach, and Shiu (2022). We now make clear the origin of this problem in a preliminary field-dependent theoretical model which confers to the recovery of $g(\Gamma) = (m_0, f_\varepsilon)$ a broader interpretation. We also briefly comment on the "leap of faith" that stems from unobservability in this case. The stochastic model in Hu, Schennach, and Shiu (2022) has two components: a measurement error model (extending Example B.1) and a non-parametric regression model (similar to Example B.3). It can be interpreted as saying that: an outcome Y is measured perfectly; this outcome is the sum of a function of some variable X^* and some purely random factor η; the explanans X^* is always measured with some error in virtue of the recording process. The authors take the stochastic model as a starting point for their analysis, but they informally provide two field-dependent theories, stated in probabilistic terms (under an implicit i.i.d. assumption – see Remark D.1), that are supposed to generate it: one from nutrition, one from economics. For completeness, we briefly introduce them:

- Nutritional model: cholesterol levels (Y) are a function of food intake (X^*) plus some purely random factor (η); cholesterol levels (Y) are perfectly measured due their recording in laboratories; observable food intake (X) is self-reported and is thus best viewed as the sum of the true food intake (X^*) and some reporting error (ε).

- Economic model: children health status (Y) is a function of household income (X^*) and some purely random factor (η); children health status (Y) is perfectly measured by "some objective measures"; observable household income (X) is obtained from tax data and is thus best viewed as the sum of the true household income (X^*) and some reporting error (ε).

The recovery of $g(\Gamma) = (m_0, f_\varepsilon)$ from the data $((x_1, y_1), \ldots, (x_n, y_n))$ would then empirically support a nutrition-specific explanation (of cholesterol levels in terms of food intakes) or an economics-specific explanation (of children health status in terms of household income). This extension of the statistical reach is paid by an unavoidable "leap of faith" that comes from all the assumptions bearing on the unobservables as made explicit in Appendix C.1. In this case, the implied constraints on the observable distribution $P_{X,Y}$ are too mild to meaningfully reject the unobservable structure from the data – empirical adequacy is not a useful guide. Most of the justifications for any specific coupling Γ have to happen at the theoretical level (from economics or nutrition in this case). As can be seen in this example, however, the stochastic model and the probabilistic assumptions for identifications are motivated by field-dependent theoretical reasons but also by practical statistical considerations. This makes the plausibility of the statistical results within the theoretical model hard to weigh in practice.

Example D.2 (Moon, Shum, and Weidner (2018))**.** We made explicit in Appendix C.2 the inverse problem behind the model in Moon, Shum, and Weidner (2018). We now make clear the origin of this problem in a preliminary economic model which confers to the recovery of $g(\Gamma) = (\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$ a broader interpretation. We also briefly comment on the "leap of faith" that stems from unobservability in this case. The stochastic model in Moon, Shum, and Weidner (2018) has a very tight connection to economic theory: it is supposed to capture aggregate level demand obtained from individual utility maximization and is a direct extension of the standard BLP model from Berry, Levinsohn, and Pakes (1995) where the "structural error" is given an interactive fixed effects formulation. For completeness, we briefly introduce the main components of the economic model which is directly stated in probabilistic terms (with distributional shift across $j = 1, 2, \ldots, J$ and $t = 1, 2, \ldots, T$ but not across $i = 1, 2, \ldots, n$ – see Remark D.1):

- Utility and mean-utility: Individual i gets some utility from buying product j in market t which is assumed to be $u_{ijt} = \delta_{jt} + \varepsilon_{ijt} + X'_{jt} v_i$. Individual utility is thus decomposable into some invariant individual preferences v_i multiplied by some observable characteristics X_{jt} of product j in market t (e.g., price), some preference shocks ε_{ijt}, and some mean-utility δ_{jt} invariant across individuals. In this case, it is further assumed that mean-utility is given by $\delta_{jt} = X'_{jt}\beta^0 + \lambda_j^{0\prime} f_t^0 + e_{jt}$, that is, it can be decomposed into a linear function of the product-market characteristics X_{jt}, some product-specific characteristics λ_j^0 multiplied by some market-specific characteristics f_t^0, and some product-market shocks e_{jt}.

- Multinomial logit choice and aggregation: Individual i buys product j in market t if it yields the highest utility among all products in market t. It is assumed that the preference shocks ε_{ijt} are i.i.d. Gumbel so that the probability that individual i chooses product j in market t is

of the multinomial logit form as given by $\pi_{jt}(\delta_t, X_t, v_i)$ where δ_t and X_t are the mean-utility and the observable characteristics when vectorized over j. It is further assumed that the individual preferences drivers v_i are identically distributed according to some distribution. The observable market share of product j in market t is then taken as the expected value of the individual choice probabilities with respect to this distribution.

The recovery of $g(\Gamma) = (\alpha^0, \beta^0, \lambda^0 f^{0\prime}, P_{e,X,Z})$ from the data $((x_{kjt}), (s_{jt}), (z_{mjt}))$ would then empirically support an economics-specific explanation for aggregate demand in terms of individual utility maximization as specified above. The extension of the statistical reach beyond information on $P_{X,S,Z}$ is paid by an unavoidable "leap of faith" that comes from all the hypotheses bearing on the unobservable structure isolated in Appendix C.2. In this case, as in Example D.1, the implied constraints from the stochastic model and the identification step on the observable distribution $P_{X,S,Z}$ are too mild to meaningfully reject the unobservable structure from the data – empirical adequacy is of little help. Most of the litigation for or against the model has to happen at the theoretical level from within economics itself. As can be seen in this example, however, the modeling choice and the identification assumptions for the problem are not all motivated by economic reasons, some follow from practical statistical considerations. This makes the plausibility of the statistical results construed as an economic explanation hard to weigh in practice.

Example D.3 (Lei and Candès (2021))**.** The Neyman–Rubin potential outcome model considered in Lei and Candès (2021) is a workhorse of causal inference. We made explicit in Appendix C.3 a few inverse problems that can be considered within this framework. For the sake of exposition, we only consider the Average Treatment Effect (ATE) one in this section and make explicit the origin of this problem in a preliminary causal theory. The main claim of this theory (directly stated in probabilistic terms under an implicit i.i.d. assumption – see Remark D.1) is that causality can be captured by some unobservable variables called potential outcomes, namely $Y(0)$ and $Y(1)$, that correspond to what the outcome Y would be if the treatment T were set or fixed to 0 and 1, respectively. It is this particular interpretation of the variables $Y(0)$ and $Y(1)$ that gives to $\tau_{\text{ATE}} = \mathbb{E}[Y(1) - Y(0)]$ its name of average treatment effect and its causal interpretation. The only constraint on $Y(1)$ and $Y(0)$ follows from a natural interpretation of potential outcomes that for each $t \in \{0, 1\}$, the variable $Y(t)$ is equal to the outcome Y whenever $T = t$. It is implicit that when $T \neq t$, the variable $Y(t)$ is unobservable. This mild clause on observability is exactly where the stochastic model at the basis of the Neyman–Rubin framework originates. In the case of the average treatment effect, the recovery of $g(\Gamma) = \tau_{\text{ATE}}$ from the data $(y_1, x_1, t_1), (y_2, x_2, t_2), \ldots, (y_n, x_n, t_n)$ would then empirically support a causal claim that the treatment T has an average causal effect on Y equal to τ_{ATE}. The cost of this causal extension of the statistical apparatus is an inevitable "leap of faith" which stems directly from the unobservability of $(Y(0), (1))$ within the inverse problem – a direct transcription of the "fundamental problem of causal inference". For the average treatment effect, this "leap of faith" comes mostly from the identification assumptions (2) and (3) of mean independence for $Y(1)$ and $Y(0)$ as stated in Appendix C.3. These assumptions cannot be empirically tested because no realizations of $(Y(0), (1))$ exist. Perfect knowledge of

the distribution of T, which happens often in practice when one is in charge of the experiment (e.g., $T \sim \text{Bernoulli}(1/2)$), would not be enough to guarantee the validity of the conditions. Only knowledge of the conditional distribution of $Y(0)$ and $Y(1)$ on T would do but it is severed from the data by definition. Put another way, any form of selection (in average) would have to be excluded ex principio for the conditions to hold. The same remarks would hold for the strong ignorability condition. This is the trade-off proper to any inverse problem here in a causal context which thus reads as: no causality without a "leap of faith". It is clear, moreover, in this case, that the constraints implied by the stochastic model and the identification assumptions on the observable distribution $P_{Y,X,T}$ are minimal – empirical adequacy is of no help here. The plausibility of the constraints on the unobservables $Y(0)$ and $Y(1)$ have to be judged based on counterfactual reasoning only.

Remark D.1. In all the problems above, we have not explicitly discussed sampling as an hypothesis in need of interpretation. It is, however, a key modeling assumption that is always running in the background of inverse problems and that plays a substantial role in the respective explanations they sustain – see Remark 4.8. It is easily seen in the previous examples that different sampling hypotheses sustain different theories by isolating different invariant parts of the joint distribution for the sample under consideration. The choice of a sampling assumption happens either when transposing a deterministic model in stochastic terms or directly when a probabilistic model is used as a starting point for the explanation. At one extreme, we have the i.i.d. hypothesis where all observations are supposed to be independent draw of the same underlying distribution – see Examples D.1 and D.3. At the other extreme, we have the unfeasible case where the joint distribution for the whole sample is not constrained. In between, we have a plethora of possibilities where only some parts of the marginal distributions are made invariant – see, for instance, Example D.2.