Ratio Mathematica Volume 51, 2024

Determining dependence among random variables across observations

Peter J. Veazie*

Abstract

It is common in applied research to analyze data from data generating processes with dependencies among random variables across observations. Such dependencies impact power calculations and standard errors. However, it is also common to mistake the structure of data for the structure of the data generating process and thereby to use inappropriate standard error estimators. The challenge is not merely to distinguish data from data generating processes but also to determine dependence. This paper discusses the problem and provides a four-step guide, with examples, for determining dependence of random variables across observations.

Keywords: clustered standard errors; data generating process; dependent random variables; hierarchical models; nested data.

2010 AMS subject classification:62-02; 62P10; 62P20; 62P25.†

* University of Rochester, Rochester NY, USA; peter_veazie@urmc.rochester.edu.

† Received on November 8, 2023. Accepted on January 15, 2024. Published on January 30, 2024. DOI:10.23755/rm.v51i0.1464. ISSN: 1592-7415. eISSN: 2282-8214.

1. Introduction

In applied research, particularly that which focuses on populations such as the social and health sciences, it is not uncommon to find studies that confuse the structure of data for the structure of the data generating process (DGP). In doing so, such work mistakes potential commonalities among the objects of measurement (e.g. students in the same school and patients in the same hospital) for dependence in the joint distributions of random variables across the observational process of the DGP. Such confusion can lead to the application of inappropriate standard error estimators. For example, see research on mortality rates among older hospitalized patients [1], readmission penalties among hospitals [2], and activities of daily living among nursing home residents [3], each of which used clustered standard errors that do not reflect the DGP. This problem is not merely due to the misreading of statistical texts, the texts themselves, intent on teaching the proper use of clustered standard errors, are sometimes misleading in how they frame the issue. For example, a recent manuscript in the Journal of Human Resources that instructs on cluster-robust inference frames the problem in terms of "where observations can be grouped into clusters" citing observational data that contain geographic location indicators such as village or state [4], which can lead to the mistaken interpretation that dependence is data specific.

To select proper standard error estimators, it is useful to determine the dependence of random variables among observations in a given study design based on the DGP. This paper provides guidance and examples for instructing students and applied researchers on how to understand a DGP to determine dependence, focusing examples on first-order linear dependence. The following is based on a simple conceptual understanding of probability spaces, a rigorous background is not required; indeed, readers without this background should be able to understand and apply the approach outlined here. A concise conceptual background on probability spaces can be found in books such as What Makes Variables Random [5], and a more comprehensive coverage can be found in books such as Probability and Measure [6] and A Probability Path [7]. The topic is also covered in discipline-specific works such as the econometric texts Topics in Advanced Economics [8] and Stochastic Limit Theory [9]. However, such texts do not provide a step-by-step guide to determining dependence, as is provided here.

2. Background

The question of dependence among random variables across observations is not one regarding the data but is instead one regarding the DGP. For example, suppose we have data from a sample of people in the United States that includes a variable indicating each person's state of residence. If the outcome of interest is strongly influenced by the state in which a person lives, and this influence varies across states, then we may reason that data from persons within the same state are likely to be more similar than data from persons of different states, and consequently the random variables across observations are dependent. This is incorrect thinking. Instead, we should ask whether the distribution of possible results from one instance of generating an observation depends on the result of another instance of generating an observation and whether the

distributions of corresponding random variables vary across the observations. For example, if we randomly sample with replacement from all persons in the United States, then each observation is independent from all others, and all of the corresponding random variables across observations are independent, regardless of whether the data contains records on multiple individuals from the same states. Why is this the case? These observations are independent because the probability of obtaining an individual in one instance of generating an observation does not change due to whom we obtain on any other instance. Alternatively, suppose we generated the data by randomly sampling states, and then within each of those states we randomly sampled two individuals with replacement. In this case, we have the classic nested DGP. A process that, as is well documented, generates dependent observations within a state cluster, since if one such observation comes from a given state, then all other observations in the cluster must come from that state as well; consequently, their corresponding random variables may be dependent as well. I used the phrase "state cluster" rather than merely using "state" to differentiate the inherent clusters (or nesting) in the DGP from the states we obtain. If we sampled states with replacement and thereby have the same distribution of sampling probabilities across states for each observation, then two different instances of sampling states may yield the same state. Such a result would produce two different state-sampled clusters within which observations are dependent and across which observations are independent, even though both clusters have the same state.

3. Determining Dependence

We can determine whether random variables are dependent in four steps based on identifying the structure of our probability space (Ω, A, P) comprised of an outcome set Ω, a sufficiently granular sigma-algebra A, and probability measure P defined to represent a DGP. The outcome set Ω comprises the set of possible elements that a DGP might obtain on which random variables are defined, in many cases this is the set of objects or individuals that a DGP could produce on which measurements can be taken. For example, Ω might be identified by a sampling frame representing the set of students in a school and corresponding random variable are the measurements of student age and math aptitude; or, perhaps Ω is the set of cars manufactured at a factory in a given year and the corresponding random variables are gas mileage and color. The sigma-algebra A is a formally structured set of subsets of Ω to which the measure P assigns probabilities associated with the DGP. For example, perhaps the DGP entails an equal probability sample from the set of objects $\Omega = \{w_1, w_2, w_3\}$ and A is the set of all subsets of Ω, i.e. $A = \{\varnothing, \{w_1\}, \{w_2\}, \{w_3\}, \{w_1, w_2\}, \{w_1, w_3\}, \{w_2, w_3\}, \{w_1, w_2, w_3\}\}$, and P assigns approximately 0.3333 to each singleton in A (e.g. $\{w_2\}$), approximately 0.6667 to each element with two objects (e.g. $\{w_1, w_2\}$), 1 to the element representing Ω (i.e. $\{w_1, w_2, w_3\}$), and 0 to the empty set, $\varnothing$. The details of the sigma-algebra are not relevant for this presentation, we will consider it sufficiently granular to allow us to assign P to any subset of outcomes required.

The four steps in determining dependence are (1) describe the DGP, (2) determine the structure of the potential outcomes (Ω), (3) determine the structure of the probability measure (P), and (4) determine the dependence of random variables.

3.1. Describe the data generating process

The first step is to describe the DGP. This can typically be done in a narrative or graphical form. Its mathematical translation is the role of the next two steps. It is, however, important to have a good understanding of the process that will be modeled in the following steps. Without it, it will be difficult to provide an accurately defined probability space that underlies analysis.

It is ambiguous to merely describe the DGP by stating "I will identify hospital patients and collect diagnosis and discharge information on them". I need to know how I am going to achieve this: I need to know the process. I should say, for example, that I will use equal probability sampling of hospitals, with replacement, from the set of all hospitals in the United States. From each sampled hospital, I will use equal probability sampling of patients, with replacement, from the set of patients within each hospital during a specified time frame.

Consider the difference between a repeated measures design investigating an intervention targeting the reduction of systolic blood pressure in which (1) the times associated with measurements of systolic blood pressure are considered fixed, perhaps at baseline and every subsequent 3 months for a year, and (2) a design in which the times are randomly determined. In the first case we would need to include in our description of the design the fact that systolic blood pressure will be measured at baseline, and each of 3, 6, 9, and 12 months following an intervention under study. Whereas, in the second case we would need to state that 4 measurement times are randomly determined for each observation. This distinction is important to successfully engaging the remaining steps in determining whether observations are dependent.

3.2. Determining the structure of potential outcomes.

The second step in determining whether random variables across observations are dependent is to identify the structure of the outcome set associated with the probability space we are using to represent the DGP. For example, if we simply randomly select an individual from a population and measure qualities of interest, then the outcome set (Ω) is the population of individuals from which we will sample, i.e. $\Omega = \{w: w \text{ is a member of the population being sampled}\}$. In this case, denoting the set of individuals in the population being sampled as W, we have the outcome set specified as $\Omega = \{w: w \in W\}$. The random variables associated with the measurable space (Ω, A), in which A denotes an appropriate sigma-algebra, are functions of the individuals (perhaps measurements on individuals) that compose the population W. If, instead, our DGP is to select a state and then select a resident of that state, then possible outcomes are pairs comprising a state (s) and person (w), i.e. $\Omega = \{(s, w): s \text{ is a member of a set of states and } w \text{ is a member of the set of residents who reside in the states}\}$. In this case, denoting the set of states as S and the set of residents across states as W, the outcome set is $\Omega = \{(s, w): (s, w) \in S \times W\}$. Corresponding random variables are a function of the pair; for example,

measurements on the state or individual in the pair. Notice, we can define random variables identifying the state in which the person lives on both of these outcome sets, and corresponding variables would exist in the resultant data sets. However, notwithstanding the existence of such variables in the data, the first case, with element (w), cannot represent nested sampling, whereas the second case, with element (s, w), can represent nested sampling. Further, suppose our DGP is to select a state (s), then select a person (w) from that state, and then randomly assign a treatment (t) to the selected person. In this case, the outcome set comprises all potential triples (s, w, t). For the set of possible treatments, denoted as T, the outcome set is $\Omega = \{(s, w, t): (s, w, t) \in S \times W \times T\}$. This is different from a DGP that selects an individual w from outcome set W and records the state within which that individual resides and her current treatment. Each DGP described above could result in data with variables that indicate the person, state of residents, and treatment; yet dependence of the random variables can be different.

How do we determine the structure of the outcome set? For the specific cases in this paper, in which we are structuring elements of a probability space used to represent a DGP, we can start by considering an arbitrary observation. Specifically, we consider how the process by which we generate the data incorporate degrees of freedom into the selection of potential outcomes. For example, if an observation entails sampling a state from the set of three states $S = \{s_1, s_2, s_3\}$ in which no element has a probability equal to 1, then the observation has a degree of freedom in determining the state that will be obtained. However, if it is determined a priori that the observation will be from state s_1, then if we wish to frame it in terms of a probability space, we are sampling from the outcome set $S = \{s_1\}$, which being a singleton, has probability 1 of sampling s_1, or more generally sampling from the set $S = \{s_1, \ldots s_j, \ldots s_J\}$ in which some element j has a probability of 1: $P(s_j) = 1$. In this case, the observation has no degree of freedom regarding state. We can exclude the set of states in the definition of the outcomes set; however, including the set can help track the fact that the state is pre-determined, which I will do in the examples below; but, this requires remembering that the probability of selecting this component of the outcome is 1.

The preceding paragraph showed how a potential component of the outcome set can be eliminated from the structure of the outcome set, specifically by identifying that there are no degrees of freedom allowed by this component in determining the outcome. Another consideration is to determine whether to expand the structure to include a component. Consider a DGP that entails sampling an individual from the set W and then randomly assigning a treatment from the set T. Can we use the probability space (W, A, P) to represent this process? In this case, because the outcome set represents only the population W and does not represent a set of possible treatments, treatment can only enter as a function of W, i.e. as a random variable. Consequently, true to the nature of a function, each individual has only one treatment assigned to it. Absent the DGP, does such a function exist? Clearly not, the value assigned by the function depends on the concrete outcome from the DGP, after the uncertainty has been resolved, which does not model what we seek to model: the uncertainty of treatment assignment. Consequently, T must be part of the outcome set, thereby allowing treatment to have a

probability assigned to it, and our probability space is $(W \times T, A, P)$. Of course, if T is a singleton or one of the elements of T has probability equal to 1 (i.e. we are determined to give whomever we select a specific treatment), then we can eliminate it from the structure as was done above regarding states S and treat it as a random variable associated with W. Another example is that of repeated measures to be recorded at predetermined times (e.g. a fixed pretreatment baseline, 6 months post treatment, and 12-month post treatment observations in a clinical research setting). Since the three measurement times are set prior to the DGP and determined by it, they can be excluded from the outcome set and treated as a random variable, or if included, they must be associated with probability equal to 1.

When considering dependence between two observations, we need to structure the elements of the outcome set for the pair of observations. If the first observation has outcome set Ω_1 and the second observation has outcome set Ω_2, then the corresponding product space has the outcome set of $\Omega_1 \times \Omega_2$. If both have the same outcome set, Ω, then the outcome set for the pair of observations is $\Omega \times \Omega$, denoted also as Ω^2. What is important in this step is to consider the arbitrary elements of each component in terms of whether they share elements. For example, suppose we are sampling states from the United States (denoted as the set S) and sampling residents from within states (denoted as the set W) the outcome set for each observation is $\Omega = \{(s, w): s \in S \text{ and } w \in W\}$. Consequently, we might indicate the outcome for the first observation as (s, w) and the second observation's outcome as (s', w'). The arbitrary element of the product set is then $((s, w), (s', w'))$. The question is whether the observations share elements. For example, if this is a nested DGP in which we sample a state s and then sample two residents from within the selected state, then two observations within a state cluster must necessarily have the same state (i.e., $s = s'$) and the arbitrary element of the product is $((s, w), (s, w'))$, whereas if we are considering two observations from different states, then the arbitrary element remains $((s, w), (s', w'))$ because the DGP does not restrict either the state or the resident to be the same across observations. It is important to identify shared elements in the outcome set across observations because this information will be used to calculate the quantities of interest in the following steps.

3.3. Determining the structure of the probability measure on the outcomes set.

In this step, we determine how the DGP imposes structure on the probability of obtaining elements from the outcome set. For example, consider an outcome set comprised of individuals (w), the probability is simply $P(w)$ for obtaining each individual w (note that the probability may be different for each w). For the state and resident outcome set it is $P(s,w) = P(w|\, s) \cdot P(s)$ for each pair. If, in this case, we were to randomly select a state, and then randomly select a person from the full population rather than from within the selected state (which seems an absurd thing to do, but suppose we did so anyway), then selection of state and person would be independent of each other, and we would structure the probability as the product of marginal probabilities: $P(s,w) = P(w) \cdot P(s)$. For the state, resident, and treatment outcome set, the probability of each possible outcome can be $P(s,w,t) = P(w|\, s, t) \cdot P(t|s) \cdot P(s)$. If, in this

case, the individual is independent of treatment within state, then $P(w|\, s,\, t) = P(w|\, s)$, and we further structure the probability as $P(s,w,t) = P(w|\, s)\cdot P(t|s)\cdot\ P(s)$. If the probability of treatment was the same across state, then treatment is independent of state, $P(t|s) = P(t)$, and we have $P(s,w,t) = P(w|\, s)\cdot P(t)\cdot\ P(s)$. On the other hand, if each person had their own probability of treatment assignment that is not dependent on the state, then it would be better to specify $P(s,\, w,\, t) = P(t|w)\cdot P(w|s)\cdot\ P(s)$. In each of these cases, and any other we wish to consider, the structure of the probability measure is usefully dictated by how the DGP produces the elements of the outcome set.

To understand dependence between any two observations, we need to structure the probability measure of the product space associated with the observations. We are concerned with structuring the probability of each element of the joint outcome set. For example, if we are considering whether observations within each state cluster in a DGP where state residents are nested within state, then based on the preceding step above, in which we determined the observations share state and the arbitrary pair is $((s,\, w),\, (s,\, w'))$, we are concerned with $\mathrm{P}(s,\, w,\, s,\, w')$, which is $\mathrm{P}(w,\, w'|\, s,\, s)\cdot\mathrm{P}(s,\, s)$. However, given the redundancy in s, this is simply $\mathrm{P}(w,\, w'|\, s)\cdot\mathrm{P}(s)$. Now, if the DGP is one in which the residents are sampled with replacement independently within state, then $\mathrm{P}(w,\, w'|\, s)$ is the product $\mathrm{P}(w|\, s)\cdot\ \mathrm{P}(w'|\, s)$, and overall probability is $\mathrm{P}(s,\, w,\, s,\, w') = \mathrm{P}(w|\, s)\cdot\ \mathrm{P}(w'|\, s)\cdot\ \mathrm{P}(s)$. We have therefore structured the joint probability of a pair of observations in terms of the DGP and their unique elements (in this example, in terms of s, w, and w').

3.4. Determine dependence of random variables between observations.

In this last step, we write random variables, as defined on the underlying probability space $(\Omega,\, A,\, P)$, in terms of the elements of the outcome set Ω and not in terms of the derived probability measure of the random variable itself. In other words, we write a random variable Y as $Y(w)$ in which w is an arbitrary element of Ω. For example, if $\Omega = S \times W$, we write an associated random variable Y as $Y(s,\, w)$ for which $s \in S$ and $w \in W$, and we consider probabilities associated with $(s,\, w)$ in a sufficient sigma-algebra (e.g. the power set of Ω).

For ease of presentation, I will consider only linear dependence (i.e. covariance); however, the same method can be used for higher order dependence, by substituting the appropriate moment, or calculation, for the covariance in the examples below. Consequently, because we are interested in the covariance, we require the expected value of the random variable in terms of the underlying probability space. For example, continuing the preceding, we insert the probability structure determined in step two above, i.e. $P(s,w) = P(w|\, s)\cdot P(s)$, into

$$E(Y_i) = \sum_s \sum_w Y(s,w)\cdot P(s,w)$$

and evaluate, say,

$$E(Y_i) = \sum_s \sum_w Y(s,w)\cdot P(w\,|\,s)\cdot P(s)\,.$$

With the expected value in hand, we can evaluate the moment equation, or equations, that represent dependents.

Continuing our example, the covariance between random variables of two observations, Y_i and Y_j, would be based on the corresponding two observations (s, w) and (s', w'):

$$Cov(Y_i, Y_j) = \sum_{s} \sum_{w} \sum_{s'} \sum_{w'} \left(Y_i(s,w) - E(Y_i)\right) \cdot \left(Y_j(s',w') - E(Y_j)\right) \cdot P(s,w,s',w') \, .$$

The required number of summations, the order of the summations, and the structure of the probability P is determined by the preceding steps applied to specific problem being addressed. Carefully reducing the covariance will provide the representation of dependence we set out to determine. Although here I have presented the steps in terms of simple random variables, Y, we can extend this analysis to functions of random variables. Details of these steps will become evident in the following examples.

4. Examples

4.1. Example 1. Standard nested sampling design

Suppose we have data on residents from a set of states in the United States of America. The data are from multiple observations from within any given state in the sample. Are random variables across observations within the same state dependent? As mentioned above, the answer depends on the nature of the DGP. To determine dependence, consider the four steps outlined above.

Describe the DGP. We randomly sample states with equal probability and replacement as the first stage of sampling, and we keep track of which state is obtained from each sample. Since we are sampling with replacement, it is possible that we obtain the same state more than once, in which case we need to denote these as separate primary sampling units. Next, we randomly sample residents with replacement from within each of the states obtained from the first stage of sampling and measure our quantities of interest.

Determine the structure of the outcome set. This DGP obtains states and residents, consequently the possible components of the outcome set are the set of states (S) from which we obtained the primary sampling units and the set of residents across the states (W). Should we include the set of states S as a component of the outcome set? To answer this, we consider whether there is any degree of freedom in determining the state that could be obtained for a given observation? Indeed, there is. For any given observation, the observation is not fixed a priori to any given state but could well have obtained any one of the states. Consequently, we must include S as a component of the outcome set.

Should we include the set of residents W as a component of the outcome set? Similar to the selection of states, the resident for any observations could well have been any one of the residents in the state that was obtained in the first phase of sampling. Consequently, as there is more than one resident in each state, the sampling of resident within state provides a degree of freedom in determining the outcome and the set of

residents must also be included in the outcome set. The outcome set is therefore $\Omega = S \times W$ with arbitrary element $\{(s, w)\}$.

For two observations from within the same primary sampling unit (i.e. the same state that was obtained for the observations) the outcome set is then Ω^2 with arbitrary element $\{((s, w), (s', w'))\}$ in which (s, w) is the state and resident associated with one observation and (s', w') is the state and resident associated with the other observation. However, since in this DGP we are investigating the dependence of random variables across observations with the same primary sampling unit (i.e. same state), then s is the same as s'. The pair of observations is therefore $\{((s, w), (s, w'))\}$ denoted with each observation having the same state s. This is important to identify because in the following steps we will need to account only for the unique individual components, which in this case is s, w, and w'.

Determine the structure of the probability measure. For an individual observation, we can structure the probability of an event $\{(s, w)\}$ as the probability of obtaining a state s multiplied by the conditional probability of obtaining a resident given a that state, i.e. $P(s, w) = P(s) \cdot P(w \mid s)$, or as the probability of obtaining an individual w multiplied by the probability of obtaining a state s given obtaining that individual, i.e. $P(s, w) = P(w) \cdot P(s \mid w)$. The former, however, more naturally aligns with how we conceptualize the DGP by which we obtain a state and then obtain a resident from within that state.

For a pair of observations from the same state as the primary sampling unit, the probability of the event $\{((s, w), (s, w'))\}$ can be written as $P(s, w, s, w') = P(s \mid s, w, w') \cdot P(s, w, w')$. However, because the probability of an event conditional on that event is 1, i.e. $P(s \mid s, w, w') = 1$, we can write the probability of event $\{((s, w), (s, w'))\}$ as $P(s, w, w')$ rather than the somewhat redundant $P(s, w, s, w')$. Moreover, the probability $P(s, w, w')$ can be expressed as the probability of obtaining state s multiplied by the probability of obtaining the pair of residents w and w': $P(s, w, w') = P(s) \cdot P(w, w' \mid s)$. By the DGP, residents are selected at random with replacement, consequently the probability of obtaining residents is independent across observations, and therefore $P(w, w' \mid s) = P(w \mid s) \cdot P(w' \mid s)$. Overall, then, we structure the probability of the joint outcome event $\{((s, w), (s, w'))\}$ as $P(s, w, w') = P(s) \cdot P(w \mid s) \cdot P(w' \mid s)$.

Determine dependence. To determine dependence of random variables across observations, we need to express the random variables as a function of the outcome set. So, as mentioned above, let us consider an outcome set comprised of State (S) and Person (W), with arbitrary individual elements of $(s, w) \in S \times W$. In this case, we denote the outcome as $Y_i(s, w)$ and the expected value of the outcome as $\bar{Y}_i$. With the expected value of Y in hand, which is the same for each observation in a given state, we can determine the covariance between the Y's of the observations within state. To calculate the covariance, we need to sum over the elements of w, w', and s:

$$Cov(Y_1, Y_2) = \sum_s \sum_{w'} \sum_w \left(Y_1(s, w) - \bar{Y}_1\right) \cdot \left(Y_2(s, w') - \bar{Y}_2\right) \cdot P(w \mid s) \cdot P(w' \mid s) \cdot P(s).$$

Note, summing over w provides the expected value of Y_1 conditional on s:

$$E(Y_1 \mid s) = \sum_{w} \left(Y_1(s,w) \cdot P(w \mid s) \right).$$

And, summing over w' provides the expected value of Y_2 conditional on s:

$$E(Y_2 \mid s) = \sum_{w'} \left(Y_2(s,w') \cdot P(w' \mid s) \right).$$

Consequently, we get

$$Cov(Y_1, Y_2) = \sum_{s} \left(E(Y_1 \mid s) - \bar{Y}_1 \right) \cdot \left(E(Y_2 \mid s) - \bar{Y}_2 \right) \cdot P(s).$$

Since the observations are independent within state, $\left(E(Y_1 \mid s) - \bar{Y}_1 \right) = \left(E(Y_2 \mid s) - \bar{Y}_2 \right)$ and the covariance between observations is the variance of the conditional expectation of Y across the states for a given observation. Remember, the correlation between Y_1 and Y_2 is the covariance divided by the product of the standard deviations for each of Y_1 and Y_2. Because the standard deviations for each of Y_1 and Y_2 are identical, this product is simply the variance of either one. Consequently, the correlation is the between-state variance divided by the total variance, which is the intraclass correlation defined for a simple model with a state-level random intercept [10]. This is the expected result for the classic nested DGP.

4.2. Example 2. Fixed primary sampling units with unit-level random treatment assignment

For a more complicated example, suppose we have data on residents from each of the 50 states in the United States of America. Moreover, suppose each state has its own probability of adopting a policy that will affect everyone in the state. Are random variables associated with a pair of observation from within the same state dependent?

Describe the DGP. We randomly obtain a sample of residents, with replacement, from each of the 50 states in the United States of America. Each state has its own probability of adopting a policy, which we can generally consider as having a probability of being subject to a treatment or intervention. For any given state, all residents will either be subject to the policy or not subject to the policy; this is a state-level treatment.

Determine the structure of the outcome set. For this DGP, the outcome set for each observation has three possible components: states, individuals, and treatments (i.e., policies in this case). Should we include a set of states that underlie each observation as a component of the outcome set? Notice that since each observation comes from a fixed pre-determined state, the set of possible states that underlies an arbitrary observation i from state s is $S_i = \{s\}$ for which $P(s) = 1$. Consequently, we have the option of dropping S from the outcome set specification. Because, within each state, the DGP will randomly select a resident, each observation has the potential of obtaining one of many possible residents (assuming each state has more than one resident). Consequently, we must include the population of residents for each observation i, W_i, as

part of $W = \bigcup_i W_i$, as a component of the outcome set. Similarly, because each state has a probability of either adopting the policy or not adopting the policy (i.e. being subject to the treatment), then the DGP has a degree of freedom associated with treatment, T = {adopts policy, does not adopt policy}; therefore, T must be a component of the outcome set. Consequently, we could use the outcome set specified as $\Omega = W \times T$. However, to keep track of the fact that observations come from fixed states, we can use $\Omega = S \times W \times T$ as our outcome set and remember that $P(s) = 1$ for the state s underlying each observation.

Regarding joint observations, consider two from the same state: the first observation will select an element (s, w, t), the second will select an element (s', w', t'). The DGP is such that they not only have the same state, but they will thereby have the same treatment (i.e. since t is a state-level assignment, it will be the same for all persons in the state). Therefore, across the pair of observations from the same selected state, $s = s'$ and $t = t'$, which I will denote simply as s and t. Our two elements will then be (s, w, t) and (s, w', t) in which both elements share state and treatment. Consequently, when we determine dependence, we will sum over s, w, w', and t.

Determine the structure of the probability measure. Because in this example we are not sampling states but are using a fixed set of states, the probability of obtaining an individual depends on the state s from which we are sampling, but it does not depend on the treatment t. Consequently, $P(w|\, s, t) = P(w|\, s)$. Similarly, treatment depends on the state (each state can have their own probability of adopting the treatment/policy), but treatment does not depend on the individual, which is to say all residents of the state will either be subject to the treatment or not. So, we structure the probability of obtaining an element (s, w, t) as the probability of obtaining person w given state s, multiplied by the probability of being assigned treatment t given state s, multiplied by the probability of obtaining state s:

$$P(s, w, t) = P(w \mid s) \cdot P(t \mid s) \cdot P(s).$$

Also, for our DGP, conditional on state, the selection of individuals within state are independent, and consequently regarding the joint distribution for observations (s, w, t) and (s, w', t) the probability of the pair of individuals conditional on state is

$$P(w, w' \mid s) = P(w \mid s) \cdot P(w' \mid s).$$

Therefore, the overall probability of both observations can be structured as

$$P(w, w', s, t) = P(w \mid s) \cdot P(w' \mid s) \cdot P(t \mid s) \cdot P(s).$$

Determine dependence. In this example, let us represent an outcome Y as a linear function of policy or treatment T, such that for arbitrary observation i they are related as

$$Y_i = \alpha + \beta \cdot T_i + \varepsilon_i.$$

To determine dependence of random variables across observations, we need to express these random variables as a function of the outcome set. So, as mentioned

above, let us consider an outcome set Ω comprised of State (*S*), Person (*W*), and Treatment (*T*), with arbitrary individual elements of $(s, w, t) \in S \times W \times T$. In this case, we have

$$Y_i(s,w,t) = \alpha + \beta \cdot T_i(s,w,t) + \varepsilon_i(s,w,t).$$

Because the covariance requires knowing the expected value of the random variable itself, we need to first calculate the expected value by summing across the components of the outcome set:

$$E(Y_i) = \sum_s \sum_t \sum_w \left(\alpha + \beta \cdot T_i(s,w,t) + \varepsilon_i(s,w,t)\right) \cdot P(w \mid s,t) \cdot P(t \mid s) \cdot P(s).$$

In terms of the reduced probability structure, the expected value of Y is then

$$E(Y_i) = \sum_s \sum_t \sum_w \left(\alpha + \beta \cdot T_i(s,w,t) + \varepsilon_i(s,w,t)\right) \cdot P(w \mid s) \cdot P(t \mid s) \cdot P(s),$$

which, after distributing the summation over individuals, *w*, can be re-expressed as

$$E(Y_i) = \sum_s \sum_t \left(\alpha + \beta \cdot \sum_w T_i(s,w,t) \cdot P(w \mid s) + \sum_w \varepsilon_i(s,w,t) \cdot P(w \mid s) \right) \cdot P(t \mid s) \cdot P(s).$$

Because for this DGP $P(w| s) = P(w| s, t)$, each of the summations with respect to *w* provide expected values conditional on *s* and *t* , i.e.

$$E(T_i \mid s,t) = \sum_w T_i(s,w,t) \cdot P(w \mid s)$$

and

$$E(\varepsilon_i \mid s,t) = \sum_w \varepsilon_i(s,w,t) \cdot P(w \mid s).$$

Therefore, the expected value of Y_i is

$$E(Y_i) = \sum_s \sum_t \left(\alpha + \beta \cdot E(T_i \mid s,t) + E(\varepsilon_i \mid s,t)\right) \cdot P(t \mid s) \cdot P(s).$$

Distributing the summations over *t* yields

$$E(Y_i) = \sum_s \left(\alpha + \beta \cdot \sum_t E(T_i \mid s,t) \cdot P(t \mid s) + \sum_t E(\varepsilon_i \mid s,t) \cdot P(t \mid s) \right) \cdot P(s).$$

Notice that for a given state *s*, the expected value of *T* given $t = 1$ is just 1, and the expected value of *T* given $t = 0$ is just 0, consequently

$$\sum_t E(T_i \mid s,t) \cdot P(t \mid s) = P(t = 1 \mid s).$$

Therefore, the expected value of *Y* is

$$E(Y_i) = \sum_s \left(\alpha + \beta \cdot P(t=1 \mid s) + E(\varepsilon_i \mid s)\right) \cdot P(s).$$

For simplicity of presentation, let us adopt the common assumption that the expectation of the error term is 0 within state, $E(\varepsilon_i \mid s) = 0$, for all s ; therefore, the expected value of Y is

$$E(Y_i) = \sum_s \left(\alpha + \beta \cdot P(t=1 \mid s)\right) \cdot P(s).$$

For this DGP, remember that each observation comes from a given state with probability equal to 1: i.e. we determine to sample an individual from a particular state in each DGP instance, so the probability of s is 1 for the s from which the observation is sampled and 0 for all other states. Therefore, the summation across states reduces to the single summand associated with the state underlying the observation, and the expected value for all observations i from predetermined state s is

$$E(Y_i) = \alpha + \beta \cdot p_s,$$

in which p_s denotes $P(T=1 \mid s)$, representing the state-specific probability of treatment.

With the expected value of Y in hand, which is the same for each observation in a given state, we can now determine the covariance between the Y's of the observations within state. To calculate the covariance, we need to sum over the elements of w, w', t, and s:

$$Cov(Y_1, Y_2) = \sum_s \sum_t \sum_{w'} \sum_w \left(\alpha + \beta \cdot T_1(s,w,t) + \varepsilon_1(s,w,t) - (\alpha + \beta \cdot p_s)\right)$$
$$\cdot \left(\alpha + \beta \cdot T_2(s,w',t) + \varepsilon_2(s,w',t) - (\alpha + \beta \cdot p_s)\right) \cdot P(w' \mid s) \cdot P(w \mid s) \cdot P(t \mid s) \cdot P(s).$$

Distributing the summations of the w's and again assuming the errors have expectation 0 (we don't need to do this, but it simplifies the presentation here), we get, noting that the α's cancel,

$$Cov(Y_1, Y_2) = \sum_s \sum_t \left(\beta \cdot E(T_1 \mid s,t) - \beta \cdot p_s\right) \cdot \left(\beta \cdot E(T_2 \mid s,t) - \beta \cdot p_s\right) \cdot P(t \mid s) \cdot P(s).$$

Expanding the summation over treatment $t \in \{0, 1\}$, and remembering that the expected value of T conditional on t is just the value of t, the covariance is

$$Cov(Y_1, Y_2) = \sum_s \left(\beta \cdot 1 - \beta \cdot p_s\right) \cdot \left(\beta \cdot 1 - \beta \cdot p_s\right) \cdot p_s + \left(0 - \beta \cdot p_s\right) \cdot \left(0 - \beta \cdot p_s\right) \cdot (1 - p_s) \cdot P(s)$$

Bringing β out from within the summation and expanding the terms under the summation yields

$$Cov(Y_1, Y_2) = \beta^2 \sum_s (1 - p_s) \cdot (1 - p_s) \cdot p_s + (0 - p_s) \cdot (0 - p_s) \cdot (1 - p_s) \cdot P(s),$$

and, again noting that, for this DGP, $P(s) = 1$ for the s underlying the observation and $P(s) = 0$ for all other states, the summation reduces to only the one summand associated with the s from which the observation was taken. Consequently, we have

$$Cov(Y_1, Y_2) = \beta^2 \cdot \left[(1 - p_s) \cdot (1 - p_s) \cdot p_s + (0 - p_s) \cdot (0 - p_s) \cdot (1 - p_s) \right],$$

which can be expressed simply as

$$Cov(Y_1, Y_2) = \beta^2 \cdot (1 - p_s) \cdot p_s .$$

The covariance of the Y's between observations within state is equal to the squared treatment effect multiplied by the variance of the state-level indicator for treatment:

$$Cov(Y_1, Y_2) = \beta^2 \cdot Var(T \mid s).$$

Consequently, the random variables are dependent unless there is no treatment effect (i.e., $\beta = 0$), or if the treatment (i.e., in this case state policy) is determined (i.e., var(T | s) = 0).

4.3. Example 3. Fixed primary units with individual-level random treatment assignment

Suppose the DGP was one in which we had a fixed set of states, as in example 2, and independent samples of individuals from within each state, also as in example 2, but our state-level probability of treatment is applied to each individual. In other words, suppose that instead of everyone in the state either receives treatment or does not receive treatment, each person "flips the state-level coin" to determine their treatment status—i.e. each person has the same probability of treatment, but this will generate some people getting treatment and others not getting treatment according to that probability. Does this imply different results from those found for example 2?

Describe the DGP. We obtain a sample of individuals, with replacement, from each state in a fixed set of states. Treatment is assigned to each individual according to a state-specific probability: i.e., individuals in the same state have the same probability of treatment, but whether they obtain treatment is independent of whether other individuals in the state obtain treatment.

Determine the structure of the outcome set. As in example 2, there are three possible components of the outcome set: states, individuals, and treatments. Given that each observation comes from a fixed, predetermined, state, we can choose not to include the set of states (S) in the outcome set, or we can include it, remembering that the probability of the state for each observation is 1 for the state from which that observation comes. Assuming each state has more than one resident, and as individuals are being sampled such that no individual has a probability equal to 1 of being obtained, there is a degree of freedom in determining the individual component of the observation: the population of individuals (W) must be included in the outcome set. And, each observation may or may not be subject to treatment according to the treatment selection probability; consequently, treatment provides a degree of freedom in specifying the observation, and treatment (T) must be included in the outcome set. The outcome set is therefore the same as that of example 2: $\Omega = S \times W \times T$.

Determining dependence among random variables across observations

Unlike example 2 our observations from the same state will be (*s, w, t*) and (*s, w′, t′*). The observations share a state but do not necessarily share treatments. Consequently, we allow treatment to be different across observations and *t* is not necessarily the same as *t′*. Therefore, when we calculate the covariance we sum over *t′* as well as *w, w′, s,* and *t* used in the preceding example.

Determine the structure of the probability measure. For arbitrary observation (*s, w, t*), the probability can be expressed as $P(s,w,t) = P(w \mid s,t) \cdot P(t \mid s) \cdot P(s)$. However, the selection probability of individual *w* from the state *s* does not depend on the treatment, consequently we can write $P(w \mid s,t)$ as $P(w \mid s)$. Also, because for each observation the state is fixed, we must remember that the probability of the state is 1 for the state underlying the specific observation.

The joint probability of observations (*s, w, t*) and (*s, w′, t′*) can be expressed as

$$P(w,w',s,t,t') = P(w,w' \mid t,t',s) \cdot P(t,t' \mid s) \cdot P(s).$$

However, again, because the selection of individuals is independent of treatment for a given state, $P(w,w' \mid t,t',s)$ can be written as $P(w,w' \mid s)$, and because individuals are selected with replacement, individual selection is independent given state and $P(w,w' \mid s)$ can be written as the product of probabilities $P(w \mid s) \cdot P(w' \mid s)$. Because the probability of treatment for one observation does not depend on the resulting treatment of another observation, $P(t,t' \mid s)$ can be rewritten as the product $P(t \mid s) \cdot P(t' \mid s)$. Consequently, the joint probability is

$$P(w,w',s,t,t') = P(w \mid s) \cdot P(w' \mid s) \cdot P(t \mid s) \cdot P(t' \mid s) \cdot P(s).$$

Determine dependence. The covariance between two observations within the same state is expressed as follows, noting that the expected value of *Y* is the same as in example 2:

$$\begin{aligned} Cov(Y_1,Y_2) = \sum_s \sum_{t'} \sum_t \sum_{w'} \sum_w & \left(\alpha + \beta \cdot T_1(s,w,t) + \varepsilon_1(s,w,t) - (\alpha + \beta \cdot p_s)\right) \\ & \cdot \left(\alpha + \beta \cdot T_2(s,w',t') + \varepsilon_2(s,w',t') - (\alpha + \beta \cdot p_s)\right) \cdot P(w' \mid s) \\ & \cdot P(w \mid s) \cdot P(t \mid s) \cdot P(t' \mid s) \cdot P(s). \end{aligned}$$

First, considering the summation over *w*, note that

$$\begin{aligned} \sum_w \left(\alpha + \beta \cdot T_1(s,w,t) + \varepsilon_1(s,w,t) - (\alpha + \beta \cdot p_s)\right) \cdot P(w \mid s) = & \beta \cdot E(T_1(s,w,t) \mid s,t) \\ & + E(\varepsilon_1(s,w,t) \mid s,t) - \beta \cdot p_s \end{aligned}$$

and for simplicity if we take the conditional expectation of the error to be 0, the summation is

$$\beta \cdot \left(E(T_1(s,w,t) \mid s,t) - p_s\right).$$

And similarly, the summation over *w′* is

$$\beta \cdot \left(E(T_1(s, w', t') \mid s, t') - p_s \right).$$

The covariance is the summation over state and treatments of these expected values multiplied by the corresponding probabilities. The covariance is therefore

$$Cov(Y_1, Y_2) = \beta^2 \cdot \sum_s \sum_t \sum_{t'} \Big[\left(E(T_1(s, w, t) \mid s, t) - p_s \right) \cdot \left(E(T_1(s, w', t') \mid s, t') - p_s \right) \cdot P(t \mid s) \cdot P(t' \mid s) \cdot P(s) \Big]$$

However, the summation over treatment t is equal to 0:

$$\sum_t \left(E(T_1(s, w, t) \mid s, t) - p_s \right) \cdot P(t \mid s) = (1 - p_s) \cdot p_s + (0 - p_s) \cdot (1 - p_s) = (1 - p_s) \cdot (p_s - p_s) = 0$$

And, similarly, the summation over treatment t' is equal to 0:

$$\sum_{t'} \left(E(T_1(s, w, t') \mid s, t') - p_s \right) \cdot P(t' \mid s) = 0 .$$

The covariance of Y_1 and Y_2 is therefore

$$Cov(Y_1, Y_2) = \beta^2 \cdot \sum_s (0) \cdot (0) \cdot P(s) .$$

Which is clearly 0. However, to complete the derivation in light of the DGP, we should note that $P(s)$ is 1 for the state from which the observations come and $P(s)$ is 0 for states from which the observations do not come. Consequently,

$$Cov(Y_1, Y_2) = \beta^2 \cdot (0) \cdot (0) = 0 .$$

From this, we conclude that the random variables of observations from within states in this DGP are independent.

In summary, for a fixed set of states (i.e., no sampling of states), if each state has a probability of treating everyone, as in example 2, then there is dependence of the random variables, if treatment has an effect and treatment is not determined. The standard errors should account for this state-level "clustering" effect. However, if each state has a probability of treatment that applies to each individual of the state (i.e. each individual obtains treatment with that probability), as in example 3, then the observations, and consequently the Y's, are independent. In this case, the standard errors should not account for a state-level "clustering" effect. This statement is true regardless of values associated with statistics such as intraclass correlations, which are trumped by the independence in the structure of the DGP.

4.4. Example 4. Random primary units with individual-level random treatment assignment

In this example, we determine the covariance between Y_i and Y_j for two arbitrary observations (i, j) within an arbitrary nursing home; however, unlike the preceding

examples in which we used fixed primary units, in this case we will randomly sample the nursing homes.

Describe the DGP. We randomly sample nursing homes with replacement from all nursing homes in the United States and then randomly sample residents with replacement from within each nursing home that is obtained. Further, each nursing home has a probability of treatment for its residents and each resident is provided treatment according to the nursing home's probability of treatment, in other words, each resident "flips their nursing home treatment coin" to determine treatment status. Notice, this is similar to example 3, except that in this case we are sampling nursing homes rather than using predetermined nursing homes, as we did using predetermined states in example 3.

Structure of the outcome set. The potential components of this DGP may at first appear to be the set of nursing homes (H), the set of nursing home residents (R), and the set of treatments (T). If these were our potential components, which of these must be part of the outcome set? To help identify the components, we can first ask whether each component has any degrees of freedom available for an arbitrary observation: is the set of nursing homes with non-zero probabilities available for an arbitrary observation a singleton? If so, then H need not be part of the formal outcome set; although, as shown in the preceding examples, it can be, but one must be careful to assign the appropriate probability of 1. However, in this example the DGP is one of randomly selecting a nursing home from a large set of nursing homes; consequently, the set of nursing homes contains multiple possible nursing homes and is thereby not a singleton. H must be a component of the outcome set. Assuming nursing homes tend to have more than one resident, the random sampling of a resident from a selected nursing home means there is more than one resident with a nonzero probability of being selected; R must be a component of the outcome set. Similarly, each resident has a nonzero probability of getting the treatment and a nonzero probability of not getting the treatment: T must be a component of the outcomes set. The outcome set must therefore be $\Omega = H \times R \times T$ with arbitrary element of (h, r, t).

Since our question is to determine the dependence of two observations from the same nursing home cluster, we need to consider the structure of the joint probability of $\{(h, r, t), (h', r', t')\}$. Here again, we need to make some reductions. First, we must consider whether these observations share any components. Since both observations will come from the same nursing home, i.e. $h = h'$, and we can just use one symbol in both observations. If the DGP assigned the same treatment to all residents of a nursing home, then it would be that $t = t'$; however, for this DGP, as in example 3, each resident "flips the coin" and therefore each resident can have a different treatment. Treatments must remain with different denotations. Our arbitrary pair of observations is therefore $\{(h, r, t), (h, r', t')\}$. Remember, the importance of this reduction is to identify the indices we will need to sum over when evaluating the covariance of random variables between observations. In this case, we must sum over h, r, t, r', and t'.

Structure of the probability measure. How can we best represent $P(h, r, t)$? Mathematically we can decompose this probability many ways. However, there is a sense of hierarchy, or nesting, here. In such cases it can, though perhaps not always, be best to work from inside to outside. Rather than express the probability as $P((h, r, t)) = P(h \mid r, t) \cdot P(r \mid t) \cdot P(t)$, it seems more interpretable as $P((h, r, t)) = P(t \mid r, h) \cdot P(r \mid h) \cdot P(h)$ in which we think of selecting a nursing home, $P(h)$, then obtaining a resident given a nursing home, $P(r \mid h)$, and finally treatment assignment given the resident within a nursing home, $P(t \mid r, h)$.

Next, can we simplify this probability specification? Well, $P(h)$ is as simple as it gets. But, can we simplify $P(r \mid h)$? If the probability of obtaining a resident did not depend on which nursing home was selected, then we could reduce this to merely $P(r)$. However, since only residents of the selected nursing home have nonzero probability of selection for an observation, clearly the probability a resident is selected depends on whether the resident's nursing home is selected, and we cannot reduce $P(r \mid h)$. What about $P(t \mid r, h)$? In this case, the DGP is one in which each nursing home has its own probability of treatment, and this treatment probability does not depend on the resident once the nursing home is identified. Since $P(t \mid r, h)$ is independent of r but not of h, we can simplify it to $P(t \mid h)$. Consequently, for a given observation, we have $P((h, r, t)) = P(t \mid h) \cdot P(r \mid h) \cdot P(h)$.

The full expression of the probability for a pair of observations from an arbitrary nursing home, attending to the hierarchical structure of sampling, is

$$P(t, t', r, r', h) = P(t \mid t', r, r', h) \cdot P(t' \mid r, r', h) \cdot P(r \mid r', h) \cdot P(r' \mid h) \cdot P(h).$$

Can this be reduced? Well, as previously stated the probability of treatment, given nursing home, is independent of selected resident: consequently $P(t \mid t', r, r', h) = P(t \mid t', h)$ and $P(t' \mid r, r', h) = P(t' \mid h)$. And, the probability of one observation's treatment does not depend on the treatment provided to the other observation: consequently, $P(t \mid t', h) = P(t \mid h)$. Also, conditional on nursing home, the probability of obtaining one resident does not depend on the resident obtained in the other observation: $P(r \mid r', h) = P(r \mid h)$. Consequently, the reduced expression for the probability is

$$P(t, t', r, r', h) = P(t \mid h) \cdot P(t' \mid h) \cdot P(r \mid h) \cdot P(r' \mid h) \cdot P(h).$$

Determine dependence. With the probability space defined by these outcome and probability components, assuming the sigma algebra is sufficient for the task, we can evaluate the covariance between random variables. For this example, as above, we consider covariance between random variables Y_i and Y_j expressed as a function of treatment, for $k \in \{i, j\}$:

$$Y_k(h, r, t) = \alpha + \beta \cdot T_k(h, r, t) + \varepsilon_k(h, r, t).$$

First, to evaluate the covariance, we need the expected values:

$$E(Y_i) = \sum_h \sum_t \sum_r \left(\alpha + \beta \cdot T_i(h, r, t) + \varepsilon_i(h, r, t)\right) \cdot P(r \mid h) \cdot P(t \mid h) \cdot P(h).$$

Following example 3, this reduces to

$$E(Y_i) = \sum_h \left(\alpha + \beta \cdot P(t=1 \mid h)\right) \cdot P(h) .$$

However, unlike state *s* in example 3, the probability of the nursing home *h* is not 1 for the selected observation. Consequently, unlike example 3, this expected value does not reduce to a linear function of nursing home specific probability of treatment, instead we sum over all nursing homes and obtain a linear function of the marginal probability of treatment

$$E(Y_i) = \alpha + \beta \cdot P(t=1) .$$

Which denoting the marginal probability of treatment as *p* can be expressed as

$$E(Y_i) = \alpha + \beta \cdot p .$$

In this example, the expected probability is the same across all observations regardless of nursing home.

The covariance of interest is then

$$\begin{aligned} Cov(Y_i, Y_j) = & \sum_h \sum_{t'} \sum_t \sum_{r'} \sum_r \left(\alpha + \beta \cdot T_i(h,r,t) + \varepsilon_i(h,r,t) - (\alpha + \beta \cdot p)\right) \\ & \cdot \left(\alpha + \beta \cdot T_j(h,r',t') + \varepsilon_j(h,r',t') - (\alpha + \beta \cdot p)\right) \cdot P\left(r \mid h\right) \cdot P(r' \mid h) \cdot P\left(t \mid h\right) \\ & \cdot P(t' \mid h) \cdot P\left(h\right). \end{aligned}$$

Notice there is a summation for each unique component across the two observations and the probability structure matches. Assuming the expected values of error terms are 0 across nursing homes, after summing across *r*, *r′*, *t*, and *t′* we are left with

$$Cov(Y_i, Y_j) = \beta^2 \cdot \sum_h \left(p_h - p\right)^2 \cdot P(h) ,$$

which is the squared treatment effect multiplied by the variance across nursing homes of the nursing home level treatment probability:

$$Cov(Y_i, Y_j) = \beta^2 \cdot \mathrm{var}(p_h) .$$

Consequently, the random variables Y_i and Y_j are dependent if there is a non-zero treatment effect and all nursing homes do not have the same probability of treatment.

5. Conclusion

The preceding three examples present distinct, though similar, DGPs. Examples 2 and 3 presented DGPs with fixed primary units (i.e., observations came from a priori fixed states) whereas example 4 presented a DGP with randomly sampled primary units (i.e. a random sample of nursing homes). However, in example 2 all individual within a primary unit received the same treatment according to a unit-specific treatment probability (as we would see in a cluster randomized trial [11]), whereas in examples 3 and 4, each individual received treatment according to a unit-specific treatment

probability (i.e. each individual metaphorically flipped a probability weighted coin to determine treatment status). Within-unit observations in example 2, in which all individuals in a unit receive treatment as determined by a unit-specific treatment probability, were dependent if there is a nonzero treatment effect and the probability of treatment is neither 0 nor 1. Within-unit observations in example 3, in which each individual flips a treatment coin with unit-specific probability of treatment, were independent regardless of treatment effect. And, within-unit observations in example 4, in which primary units are sampled and individuals flip a unit-specific treatment coin, were dependent if there is a nonzero treatment effect and treatment probabilities vary across primary units. These examples show that similar DGPs can produce different conclusions regarding the nature of dependence among random variables across within-unit observations.

Following the four-step strategy for identifying observational dependence presented here, it is easy to determine the dependence between observations in other DGPs. This is important to understand as it is a common mistake to use cluster standard errors when we have data that may appear to be "clustered" because we mistakenly reason data that share a group feature are more likely to be similar within group and should thereby use clustered standard errors. Consequent mistaken standard errors can lead to mistaken power calculations, confidence interval estimation, and statistical inferences in hypothesis testing. It is important to determine whether random variables across observations are dependent in virtue of analyzing the DGP—one way to accomplish this is to follow the four steps presented here.

References

[1] Y.F. Philpotts, X.Y. Ma, M.R. Anderson, M. Hua, M.R. Baldwin, Health Insurance and Disparities in Mortality among Older Survivors of Critical Illness: A Population Study, Publisher, City, 2019.

[2] M.S. Aswani, M.L. Kilgore, D.J. Becker, D.T. Redden, B. Sen, J. Blackburn, Differential Impact of Hospital and Community Factors on Medicare Readmission Penalties, Publisher, City, 2018.

[3] J. Wang, R.L. Kane, L.E. Eberly, B.A. Virnig, L.H. Chang, The Effects of Resident and Nursing Home Characteristics on Activities of Daily Living, Publisher, City, 2009.

[4] A.C. Cameron, D.L. Miller, A Practitioner's Guide to Cluster-Robust Inference, Publisher, City, 2015.

[5] P. Veazie, What makes variables random : probability for the applied researcher, CRC Press, Taylor & Francis Group, Boca Raton, 2017.

[6] P. Billingsley, Probability and Measure, 3 ed., Wiley and Sons, New York, 1995.

[7] S.I. Resnick, A Probability Path, Birkhauser, Boston, 1999.

[8] P.J. Dhrymes, Topics in Advanced Econometrics: Probability Foundations, Springer-Verlag, New York, 1989.

[9] J. Davidson, Stochastic Limit Theory: An Introduction for Econometricians, Oxford University Press, Oxford, 1994.

[10] A. Donner, A Review of Inference Procedures for the Intraclass Correlation-Coefficient in the One-Way Random Effects Model, Publisher, City, 1986.

[11] S. Puffer, D.J. Torgerson, J. Watson, Cluster randomized controlled trials, Publisher, City, 2005.