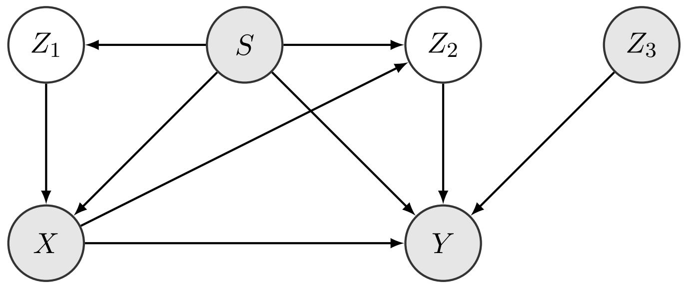

The proposed causal graph shows that the effect of probationary period length X on layoffs by tenure Y is identifiable. Note that I've grouped distinct variables into single nodes to keep notation simple. The graph suggests that all back-door paths are blocked by time, region, and occupation S. Below I will explore different strategies for estimating the causal effect.

Structural Linear Model

Assuming linear relations between all variables we get

$$y = f_Y(x, s, z_3, z_2, u_y) = \beta_1 x + \beta_2 s + \beta_3 z_3 + \beta_4 z_2 + u_y$$

With the available data we can run the regression

$$y = \beta_0 + \beta_1 x + \beta_2 s + \beta_3 z_4 + \varepsilon$$

with OLS to obtain consistent estimates of β_1. This is the traditional approach in applied economics. This approach is simple in the binary case, i.e., $x = 0$ (no probationary period) versus $x = 1$ (3 month probationary period). In the multinomial case, I would have to choose a reference probatinary period (e.g., 3 months) and have dummy variables indicating all other options with their respective causal coefficients relative to the baseline.

Following Shalizi, the approach derived from the causal graph suggests the following. For any intervention on the probationary period (i.e., setting $x = x_0$) we can get the average effect on layoffs by tenure

$$\begin{aligned}
\mathbb{E}[Y|do(x_0)] =& \mathbb{E}[f_Y(x_0, s, z_3, z_2, u_y)] \\
& \mathbb{E}[f_Y(x_0, s, z_3, z_2, u_y)|X = x_0] \\
& \sum_s \mathbb{P}(S = s_0)\mathbb{E}[f_Y(x_0, s_0, z_3, z_2, u_y)|X = x_0, S = s_0] \\
& \sum_s \mathbb{P}(S = s_0)\mathbb{E}[Y|X = x_0, S = s_0]
\end{aligned}$$

Given our linear model, the term in the expectation operator can be estimated with a regression. One could also use Monte Carlo with sample averages of layoffs by tenure within each S cell. For the outer exectation—which is with respect to $p(s)$—one can use Monte Carlo again, matching on S, (multinomial) propensity score matching, or inverse (multinomial) propensity score weighting. To get the average treatment effect between x_1 and x_0, one simply takes the difference $\mathbb{E}[Y|do(x_1)] - \mathbb{E}[Y|do(x_0)]$.

What is the differences between the simple linear regression and the matching estimators? According to Angrist and Pischke, the regression and matching estimand differ in the weights used to combine the covariate-specific effects into a single average effect. The former uses a variance-weighted average of the effects, while the latter relies on $\mathbb{P}(S = s_0)$ or how we model the propensity score. For the matching cases, there are additional details to fill in that one may not have scientific knowledge to support, e.g., how to model the score. Moreover, inference under matching is less standardized than under regressions. For this reason, I would generally prefer the simple regression approach. Although, the other approaches could be useful robustness checks.

The linear model is particularly silly for my causal question. If anything, these estimates might show the signs of causal relationships but counterfactual simulations from such a model would be laughable. The next section explores an economic model, grounded in economic theory. Estimation will be more complicated (and sensitive to the model), but to the extent that the model is correct, the stuctural parameters will be more credible.

Structural Economic Model

Denote $y(0)$ and $y(1)$ as the potential outcomes. Split the joint distribution of potential outcomes and treatment (conditional on observables) into an outcome model and an assignment mechanism.

$$p(y(0), y(1), x|s, \theta, \phi) = \underbrace{p(y(0), y(1)|s, z_3, \theta)}_{\text{Outcome}}\underbrace{p(x|y(0), y(1), s, z_3, \phi)}_{\text{Selection}}$$

Note that since the data I have is observational, it is better to refer to the assignemnt mechanism as a selection mechanism. A crucial assumption here is that the parameters of the outcome model θ are separate from those in the selection model ϕ. Moreover, given that S satisfies the backdoor criterion from X to Y, this means that selection is unconfounded. In other words, $X \perp\!\!\!\perp \{Y(0), Y(1)\}|S$ so we get

$$p(y(0), y(1), x|s, \theta, \phi) = p(y(0), y(1)|s, z_3, \theta)p(x|s, z_3, \phi)$$

For estimation puposes, the specification of the selection mechanism is not very important. This hinges on two key assumptions: 1) S satisfying the backdoor criterion; and 2) θ being separate from ϕ. I will talk more about this point in the next section. But for now I will focus on specifying the outcome model.

(i) Outcome Model

In this model, firms learn about match quality over time through Bayesian updating. Firms then solve a dynamic optimization problem where their choice variable is whether to keep the current match (retain the worker) or draw a new one (fire the worker and hire another one). Considering the empirical setting in questions, firms maximize the present discounted value of profits through match termination decisions facing a nonlinear firing cost contract and uncertainty about match quality.

Suppose that the economy consists of multiple risk-neutral, single-job, forward-looking firms $j \in \mathbb{J}$ operating over an infinite discrete time period $t \in [0, 1, 2, ...]$. Each firm enters the model at $t = 0$ with its single vacancy filled by some agent $i \in \mathbb{I}$ earning a fixed monthly wage w corresponding to the vacancy in question. Matching and wage setting considerations are ignored for this simple version of the model. Assume that the firm's only factor of production is labor and that it faces constant returns to scale. Workers assume a passive role throughout the model, while firms make firing decisions. Since firms are assumed to employ one worker at a time, the rest of this section will simply focus on a single match (i, j).

At each time period t, the firm decides on a value for the choice variable

$$d_t = \begin{cases} 1 & \text{fire the current worker and draw a new match} \\ 0 & \text{keep the current worker and reconsider next period} \end{cases}$$

The state variables that influence the binary control d_t are twofold: worker tenure (t) and beliefs about match quality (y_t). Tenure indicates how many periods the current worker has been employed at the firm in question, and therefore affects the firing cost schedule as well as the time for learning about match quality. This state variable is observed by both the firm and the econometrician. On the other hand, match quality is observed by neither the econometrician nor the firm. However, each period the firm observes signals $\xi_t \sim G(\cdot)$ with mean equal to the true match quality y_*. Based on the sequence of signals received and their priors about the distribution of match quality in the relevant population, firms make inferences about match quality at each tenure y_t.

The firm's objective is to choose the sequence $\{d_t\}_{t=0}^{\infty}$ that maximizes discounted expected profits while facing a given employment contract k with firing cost nonlinearities $C_k(w, \tau_t)$ and uncertainty about match quality determined by some density $h_t(y)$. The infinite horizon problem takes the following form:

$$\max_{\{d_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi(d_t; w, \xi_t, t)$$

where

$$\pi(d_t; w, \xi_t, t) = R(w, \xi_t) - C_k(w, t) \cdot \mathbb{1}\{d_t = 1\}$$

Both the probationary period stipulated in the employment contract k and the active legislation on firing fines introduce discontinuities in slopes and levels to the firing cost schedule $C_k(w, t)$. Meanwhile, the function $R(w, \xi_t)$ captures the revenue that the firm realizes in period t with the signal produced by the worker net of wage payments and other fixed labor costs. The timing of events within a period is such that this revenue is unaffected by the firm's firing decision. That is, within period t the timing of events are:

1. Firm receives signal $\xi_t \sim G(\cdot)$ with mean y_*
2. Revenue is generated $R(w, \xi_t)$
3. Firm updates beliefs about match quality y_t and $h_t(y)$
4. Firm makes firing decision d_t

Following the Bellman Principle of Optimality, define the choice-specific value functions $\tilde{V}(d; y, t)$ for current firing decisions as follows

$$\begin{aligned}\tilde{V}(1; y, t) =& [R(w, \xi) - C_k(w, t)] + \beta \mathbb{E}_{h_0}[V(y', 0)] \\ \tilde{V}(0; y, t) =& [R(w, \xi)] + \beta \mathbb{E}_{h_t}[V(y', t + 1)]\end{aligned}$$

where $V(y, t) = \max\{\tilde{V}(1; y, t), \tilde{V}(0; y, t)\}$ denotes the value to the firm upon choosing the optimal policy. The reason for V being a function of beliefs about match quality y_t and tenure t only holds when updated beliefs about match quality depend only on those variables (as is the case when match quality and worker signals are normally distributed). Note that the density functions over which expectations are taken are different in each choice-specific value function. This captures the notion that when a worker is fired, the random variable for match quality of some new unobserved worker has a density (h_0) that is different from that of the current worker that has been observed t periods (h_t).

(ii) Parametrization

A time period in the model will denote 3 days. This is because tenure is recorded in months (of 30 days) up to one decimal place. Thus, the monthly wages in the data map into period wages of $w/10$. Assume the simple linear revenue function $R(w, \xi) = \xi - w/10$. That is, treat signals as profit shocks and set revenue equal to the period's signal minus per period wages. Institutional knowledge discussed in previous sections gives us the form for the firing cost schedule given some probationary period length T_k and firing fines f, i.e.,

$$C_k(w,t) = \mathbb{1}\{t \leq T_k\}[0.5(T_k - t)w/10] + \mathbb{1}\{t > T_k\}[(1+b)w + f(t+10)w/10]$$

In other words, prior to the probationary period the worker is fired immediately and the firm must pay half of the remaining wages due in the probationary period. After the probationary period, the worker is notified of the separation and the firm must still pay an entire month of wages including benefits b, as well as firing fines proportional to the monthly wages accrued up to the advanced notice month $t+1$. In the case of a 3 month experience contract under present day job security legislation, $T_k = 30$ (90 days), $b = 0.358$, and $f = 1/300$.

Assume that match quality is continuous and normally distributed within each group of vacancies with the same wage, i.e., $Y_* \sim \mathcal{N}(y_0 + \frac{w}{10}, \sigma_0^2)$. The reason for varying the mean by wage is that it would be too restrictive to impose normality on productivity for the entire population. In addition, firm priors coincide with the true distribution of match quality. That is, the density h_0 is that of a normal with mean $y_0 + \frac{w}{10}$ and variance σ_0^2. Moreover, I assume that the signals about match quality are also normally distributed, i.e. $\xi_t \sim \mathcal{N}(y_*, \sigma_*^2)$, where y_* is the realized quality of a given match and σ_*^2 is constant across matches. Given that the conjugate prior of a normal likelihood is also normal, firms' beliefs over match quality have a closed form. In particular, employing a Kalman filter provides the following updated beliefs about permanent unobserved quality for some realized match at tenure t

$$y_t = \left(y_{t-1} + \frac{w}{10}\right)\left[\frac{(t-1)\sigma_0^2 + \sigma_*^2}{t\sigma_0^2 + \sigma_*^2}\right] + \left(\frac{\sigma_0^2}{t\sigma_0^2 + \sigma_*^2}\right)\xi_t$$

$$\sigma_t^2 = \frac{\sigma_0^2\sigma_*^2}{t\sigma_0^2 + \sigma_*^2}$$

This results in y_t and t fully describing the firm's knowledge about match quality—hence their role as the state variables in the model. Also note that we can now specify the density $h_t = f(y_{t+1}|y_t)$. By using the closed form for y_{t+1}, conditioning on y_t, and recalling that $\xi_{t+1} \sim \mathcal{N}(y_*, \sigma_*^2)$, we again obtain a normal $\mathcal{N}(\mu, s^2)$ where

$$\mu = \left[\frac{\sigma_0^2}{(t+1)\sigma_0^2 + \sigma_*^2}\right] y_* + \left[\frac{t\sigma_0^2 + \sigma_*^2}{(t+1)\sigma_0^2 + \sigma_*^2}\right] \left(y_t + \frac{w}{10}\right)$$
$$s^2 = \left(\frac{\sigma_0^2}{(t+1)\sigma_0^2 + \sigma_*^2}\right)^{1/2} \sigma_*^2$$

To wrap up this simple model, I fix the discounting factor β to 0.95. This means that in this simple version of the model there are only 3 parameters to estimate:

1. Two parameters determining firms' priors y_0 and σ_0^2
2. One parameter dictating the signal generating process σ_*^2

Given any value of these parameters, the firm's problem can be solved with a value function iteration (VFI) algorithm.

Open Questions

Even though my model includes Bayesian updating, I am not sure if this is the type of Bayesian model we are discussing in class. I entered this course hoping to find the link between the causal inferences we do in economics and causal graphs. My objective was to establish a workflow that would allow me to propose structural models that can be estimated with the tools we learned in the graphical models course. So far, I have not been able to figure this out. Below are the open questions that I hope to answer by the end of the semester.

1. How can my model have more variables than those proposed in the causal graph? My outcome model according to the potential outcomes framework was $p(y(0), y(1)|s, z_3, \theta)$. Mapping these to my proposed model, I would say that θ are the parameters of my model, i.e., $(y_0, \sigma_0^2, \sigma_*^2)$. The potential outcomes $y(0)$ and $y(1)$ are the tenure at separation under different probationary period lengths (modelled in C_k. Variable z_3 can be included in my model as discontinuities in firing costs at 6 months (UI) and 12 months (MM) of tenure. Meanwhile, s could be included by estiamting this model separately for every time-region-occupation. But what about wages, tenure, signals, etc. that are not in the causal graph?

2. Having time-region-occupation (S) as the variable that satisfied the backdoor criterion seems problematic. Since a collective bargaining agreement is set in a time-region-occupation, there is no variation in probationary period length within each cell. So where does the identifying variation come from? This issue applies even to the simple linear regression. This is why in economics we rely on quasi-experimental variation (e.g., the bunching in separations that should occur near the end of the probationary period), which cannot be expressed in causal graphs. Am I wrong about this? If causal graphs can depict quasi-experimental variation that would be great to learn.

3. The way an economist would solve the model I proposed is by estimating the parameters θ and then obtaining all potential outcomes through simulation. But this is not the imputation procedure we have been discussing in class. I have not mentioned a single prior, other than through the Bayesian updating portion of my model. I believe that the tools we are learning in class should allow me to do a lot more. But I still have not been able to connect the dots.