Задача 1

Принято считать, что инновации – ключ к экономическому росту, а хорошо работающая патентная система — залог наличия адекватных стимулов к инновациям. В последнее время, однако, раздается все больше голосов, утверждающих, что патенты могут не способствовать инновациям, а наоборот, препятствовать им. Действительно, наличие патента на изобретение X может препятствовать появлению новых изобретений, которые основываются на изобретении X или улучшают его – особенно, если процесс торга относительно лицензионных отчислений неэффективен. В статье 2015 года[1] экономисты Альберто Галассо и Марк Шанкерман используют эконометрический анализ, чтобы ответить на вопрос, действительно ли наличие патентов затрудняет последующие инновации. Авторы используют базу данных всех случаев патентных диспутов в апелляционном суде США. В качестве основной независимой переменной рассматривается бинарная переменная о том, был ли данный патент аннулирован судом ($Invalidated_i$). В качестве меры последующих инноваций (инноваций, которые основываются на изобретении, которое защищает данный патент) используется количество цитирований данного патента, произошедших *после* решения суда (переменная $PostCites_i$).

1. Объясните, почему простая регрессия $PostCites_i$ на $Invalidated_i$ не может дать ответа на вопрос, препятствуют ли патенты последующим инновациям.

2. Чтобы преодолеть трудности, описанные вами в пункте 1, авторы используют эмпирическую стратегию, основанную на том, что распределение патентных исков между судьями апелляционного суда осуществляется случайным образом. Объясните, почему такая стратегия действительно помогает ответить на поставленный в исследовании вопрос.

3. Авторы используют метод инструментальных переменных. Известно, что каждый иск рассматривается коллегией из трех судей, в которой решение о том, анулливовать патент или нет, принимается простым большинством голосов, причем судьи принимают решение независимо. Имеются данные о том, какая коллегия рассматривала каждый из случаев в имеющейся выборке. На основании этой информации предложите разумный инструмент для $Invalidated_i$, использование которого поможет ответить на поставленный в исследовании вопрос.

Решение задачи 1

1. Эндогенность переменной $Invalidated_i$ — например, патенты на менее ценные изобретения будут с большей вероятностью инвалидированы, и одновременно произведут меньше цитирований, а значит, оценка в такой регрессии будет занижена. В частности, если такая регрессия даст нулевую оценку β, это еще не будет значить, что патенты не препятствуют инновациям (что соответствует положительному значению β).

2. Случайное распределение судей – источник экзогенной вариации для $Invalidated_i$. Это означает, что наши данные являются частично экспериментальными, что позволит решить проблему эндогенности с помощью подходящего инструмента.

3. Разные судьи в разной степени склонны аннлировать патенты. Именно эта вариация в поведении судей должна стать основой для инструмента. По имеющимся данным можно для каждого судьи j вероятность анулляции патента p_j, а затем посчитать вероятность инвалидации для коллегии из судей a, b, c по правилу простого большинства: $P_{a,b,c} = p_a p_b p_c + p_a p_b(1-p_c) + p_a p_c(1-p_b) + p_b p_c(1-p_a)$. Для каждого патента i можно рассчитать

[1] Alberto Galasso, Mark Schankerman, Patents and Cumulative Innovation: Causal Evidence from the Courts , The Quarterly Journal of Economics, Volume 130, Issue 1, February 2015, Pages 317–369

$P_{(a,b,c)(i)}$, для которой при расчете p_a, p_b и p_c будут использованы все случаи, в которых принимал участие данный судья, *кроме случая* i – иначе наш инструмент не будет по-настоящему экзогенным. Именно это и делают авторы.

Задача 2

Является ли временной ряд $x_t = 1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}$ стационарным (u_t - белый шум)? Вычислите, если это возможно, коэффициенты автокорреляции первого, второго и третьего порядков для этого временного ряда.

Решение задачи 2

$$x_t = 1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}$$

$$1 - 1,3L + 0,4L^2 = 0$$

$$0,4L^2 - 1,3L + 1 = 0$$

$$D = b^2 - 4ac = 1,69 - 1,6 = (0,3)^2$$

$$L_{1,2} = \frac{-b \pm \sqrt{D}}{2a} = \frac{1,3 \pm 0,3}{0,8} => L_1 = 2, L_2 = 1,25$$

Оба корня $> 1 =>$ процесс стационарен.

$$V(x_t) = \gamma_0 = V(1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}) =$$

$$= (1,3)^2\gamma_0 + (0,4)^2\gamma_0 + \sigma^2 + (0,5)^2\sigma^2 + 2*1,3*0,4*cov(x_{t-1}, x_{t-2}) + 1,3*2*cov(x_{t-1}, u_t) -$$

$$-2*1,3*0,5*cov(x_{t-1}, u_{t-1}) - 2*0,4*cov(x_{t-2}, u_t) + 2*0,4*0,5*cov(x_{t-2}, u_{t-1}) =$$

$$= \{cov(x_{t-1}, u_t) = 0, cov(x_{t-2}, u_t) = 0, cov(x_{t-2}, u_{t-1}) = 0\} =$$

$$= (1,3)^2\gamma_0 + (0,4)^2\gamma_0 + 1,25\sigma^2 - 1,04\gamma_1 - 2*1,3*0,5*\sigma^2 =>$$

$$=> \gamma_0 = 1,85\gamma_0 - 1,04\gamma_1 - 0,05\sigma^2 => \gamma_0 = \frac{1,04\gamma_1 + 0,05\sigma^2}{0,85}$$

$$\gamma_1 = cov(x_t, x_{t-1}) = cov(1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}, x_{t-1}) = 1,3\gamma_0 - 0,4\gamma_1 - 0,5\sigma^2 =>$$

$$=> \gamma_1 = \frac{1,3\gamma_0 - 0,5\sigma^2}{1,4}$$

Тогда

$$\gamma_0 = \frac{1,04\frac{1,3\gamma_0 - 0,5\sigma^2}{1,4} + 0,05\sigma^2}{0,85} = \frac{0,966\gamma_0 - 0,371\sigma^2 + 0,05\sigma^2}{0,85} = 1,14\gamma_0 - 0,378\sigma^2 => \gamma_0 = 2,7\sigma^2$$

$$=> \gamma_1 = 2,89\sigma^2 =>_1 = \frac{\gamma_1}{\gamma_0} = 1,07$$

$$\gamma_2 = cov(x_t, x_{t-2}) = cov(1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}, x_{t-2}) = 1,3\gamma_1 - 0,4\gamma_0 = 2,677\sigma^2 =>$$

$$=>_2 = \frac{\gamma_2}{\gamma_0} = 0,99$$

$$\gamma_3 = cov(x_t, x_{t-3}) = cov(1,3x_{t-1} - 0,4x_{t-2} + u_t - 0,5u_{t-1}, x_{t-3}) = 1,3\gamma_2 - 0,4\gamma_1 = 2,324\sigma^2 =>$$

$$=>_3 = \frac{\gamma_3}{\gamma_0} = 0,86$$

Задача 3 [2]

Конечная цель анализа данных – улучшение качества принятия решений. В связи с этим в последнее время становятся все более популярными модели, в которых рассматривается непосредственно задача принятия решений на основе данных, без промежуточного этапа оценивания параметров.

Представим себе тестирование нового лекарства В, по результатам которого будет принято решение, в какой степени использовать его в клинической практике. Для простоты будем считать, что исход лечения бинарный: либо пациент поправится ($Y_i = 1$), либо нет ($Y_i = 0$). Пациенты, не получающие новое лекарство, будут принимать старое лекарство А, которое помогает с известной вероятностью $\mu_A = 0.6$. Эффект от нового лекарства (вероятность, с которой оно помогает), μ_B, до начала испытаний не известен; известно лишь, что $\mu_B \in [0; 1]$.

В испытаниях принимают участие N пациентов, все они принимают новое лекарство; для простоты в данной задаче ограничимся $N = 1$. В результате испытаний становится известным исход для данного пациента, Y. Если $Y = 1$, доля населения p_1 подвергнется лечению новым лекарством, если $Y = 0$, то эта доля будет равна p_0. Таким образом, принимаемым решением является пара (p_0, p_1).

Регулятор хотел бы максимизировать ожидаемый средний исход для популяции в целом, то есть величину

$$\begin{aligned}U(p_0, p_1, \mu_A, \mu_B) &= P(Y = 1)(p_1\mu_B + (1 - p_1)\mu_A) + P(Y = 0)(p_0\mu_B + (1 - p_0)\mu_A) = \\ &= \mu_B(p_1\mu_B + (1 - p_1)\mu_A) + (1 - \mu_B)(p_0\mu_B + (1 - p_0)\mu_A) = \\ &= (\mu_B - \mu_A)(p_1\mu_B + p_0(1 - \mu_B)) + \mu_A,\end{aligned}$$

однако она зависит от неизвестного параметра μ_B, и поэтому рассчитать ее не представляется возможным. В этой ситуации неопределенности популярными критериями для принятия решений являются (1) максимизация исхода в наихудшем случае; (2) минимизация *сожаления*.

1. Допустим, регулятор максимизирует исход в наихудшем случае, то есть решает задачу

   $$\max_{p_0,p_1} \min_{\mu_B} U(p_0, p_1, 0.6, \mu_B)$$

   Какие p_0 и p_1 ему следует выбрать?

2. Определим величину *сожаления* как $R(p_0, p_1, \mu_A, \mu_B) = \max\{\mu_A, \mu_B\} - U(p_0, p_1, \mu_A, \mu_B)$. Сожаление есть разница между максимальной полезностью регулятора, обладающего полной информацией о μ_B и его фактической полезностью от принятия решения (p_0, p_1). Допустим, регулятор минимизирует максимально возможное сожаление, то есть решает задачу

   $$\min_{p_0,p_1} \max_{\mu_B} R(p_0, p_1, 0.6, \mu_B)$$

   Какие p_0 и p_1 ему следует выбрать?

Решение задачи 3

1. Заметим, что если регулятор выберет $p_0 = p_1 = 0$, то его полезность будет гарантированно равна $\mu_A = 0.6$. Если же хотя бы одно из чисел положительно, то полезность регулятора может быть меньше 0.6 (при $\mu_B \in (0; 0.6)$). Таким образом, минимальная полезность регулятора максимальна при $p_0 = p_1 = 0$ – ему не следует тестировать новое лекарство вообще, в силу крайней осторожности. Это известный недостаток критерия “максимин” – он зачастую предписывает неинтересные угловые решения.

[2] По мотивам статьи "Minimax regret treatment choice with finite samples"by Jorg Stoye

2. $p_0^* = 0$, $p_1^* = 5/6$.

   Примечание: при $\mu_A < 0.5$ у нас получилось бы, что, наоборот, $p_0^* > 0$, $p_1^* = 1$ – если известное лекарство недостаточно неэффективно, имеет смысл пробовать новое, даже если результат испытаний неудачный. См. полное решение для $N > 1$ в работе Stoye (Journal of Econometrics, 2009).

Задача 4[3]

Пусть существуют $L > 2$ регионов (индекс одного региона - l), в каждом из которых живут n жителей (индекс одного жителя - i). Вы хотите оценить влияние бинарной политики $w \in \{0,1\}$ на переменную Y_{il}. Предположим, что она имеет следующую структуру:

$$Y_{il}(w) = \mu_l + \tau w + u_{il}$$

где u_{il} - независимые одинаково распределенные случайные величины для жителей и регионов, $\mathbb{E}[u_{il}] = 0$. Пусть $\sigma_u^2 := \mathbb{V}[u_{il}]$.

Государство интересуется эффектом политики τ и решает провести эксперимент. Рассматриваются два возможных дизайна. В первом случайным образом выбираются $\frac{L}{2}$ города (предполагается, что L четное), и применяется политика $W_{il} = 1$ ко всем жителям этих городов. К остальным жителям применяется политика $W_{il} = 0$. Второй дизайн отличается: в каждом городе случайным образом выбираются $\frac{n}{2}$ жителя (предполагается, что n четное), к которым применяется политика $W_{il} = 1$ (ко всем остальным применяется $W_{il} = 0$).

В обоих случаях в качестве оценки τ используется разность средних:

$$\hat{\tau} = \frac{\sum_{il} Y_{il} W_{il}}{\sum_{il} W_{il}} - \frac{\sum_{il} Y_{il}(1 - W_{il})}{\sum_{il}(1 - W_{il})}$$

1. Покажите, что $\hat{\tau}$ является несмещенной оценкой τ для каждого из дизайнов эксперимента.

2. Вычислите дисперсию $\hat{\tau}$ для каждого из дизайнов. Различаются ли они? Какой из дизайнов лучше (с точки зрения дисперсии)?

3. Является ли $\hat{\tau}$ состоятельной оценкой для обоих дизайнов, если n постоянно, а L стремится к бесконечности?

4. Является ли $\hat{\tau}$ состоятельной оценкой для обоих дизайнов, если L постоянно, а n стремится к бесконечности?

Решение задачи 4

The estimator can be written in the following way:

$$\hat{\tau} = \frac{1}{nL} \sum_{il} \gamma_{il}^{(1)} Y_{il}(1) - \frac{1}{nL} \sum_{il} \gamma_{il}^{(0)} Y_{il}(0) =$$

$$\frac{1}{L} \sum_{l} \mu_l \left(\frac{1}{n} \sum_{i} (\gamma_{il}^{(1)} - \gamma_{il}^{(0)}) \right) + \tau \frac{1}{nL} \sum_{il} \gamma_{il}^{(1)} + \frac{1}{nL} \sum_{il} (\gamma_{il}^{(1)} - \gamma_{il}^{(0)}) u_{il} \quad (1)$$

[3] По мотивам статьи “When Should You Adjust Standard Errors for Clustering?” by Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge

where the weights have the following form:

$$\gamma_{il}^{(w)} = \frac{\{W_{il} = w\}}{\frac{\sum_{jk}\{W_{jk}=w\}}{nL}} \tag{2}$$

By construction $\gamma_{il}^{(w)}$ depend only on $\{W_{il}\}_{il}$ and thus are independent of u_{il}. Also by construction we have the following:

$$\begin{aligned}\frac{1}{nL}\sum_{il}\gamma_{il}^{(w)} &= 1\\ \frac{1}{nL}\sum_{il}\{W_{jk} = w\} &= \frac{1}{2}\end{aligned} \tag{3}$$

For the second design we also have the following for every l:

$$\frac{1}{n}\sum_{i}\gamma_{il}^{(w)} = 1 \tag{4}$$

and for the first one we have the following for every l:

$$\frac{1}{n}\sum_{i}(\gamma_{il}^{(1)} - \gamma_{il}^{(0)}) = 2\{\text{the city } l \text{ was selected}\} - 1 \tag{5}$$

Define the following random variable:

$$\xi_l := \frac{1}{n}\sum_{i}(\gamma_{il}^{(1)} - \gamma_{il}^{(0)}) \tag{6}$$

It follows from the statements above that ξ_l is equal to $2\{\text{the city } l \text{ was selected}\} - 1$ for the first design and is equal to zero for the second design.

It follows that we have the following representation:

$$\hat{\tau} = \frac{1}{L}\sum_{l}\mu_l\xi_l + \tau + \frac{1}{nL}\sum_{il}(\gamma_{il}^{(1)} - \gamma_{il}^{(0)})u_{il} \tag{7}$$

As a result we can easily compute conditional expectation and variance:[4]

$$\begin{aligned}\mathbb{E}\left[\hat{\tau}|\{W_{il}\}_{il}, \{\mu_l\}_l\right] &= \frac{1}{L}\sum_{l}\mu_l\xi_l + \tau\\ \mathbb{V}\left[\hat{\tau}|\{W_{il}\}_{il}, \{\mu_l\}_l\right] &= \frac{\sigma^2\sum_{il}\left(\gamma_{il}^{(1)} - \gamma_{il}^{(0)}\right)^2}{(nL)^2} = \frac{4\sigma^2}{nL}\end{aligned} \tag{8}$$

By construction we have that $\mathbb{E}[\xi_l] = 0$; it is also straightforward to see that $\mathbb{V}[\xi_l] = 1$. Finally we can compute $\mathbb{E}[\xi_l\xi_k]$:[5]

$$\mathbb{E}[\xi_l\xi_k] = \mathbb{E}[\xi_l\mathbb{E}[\xi_k|\xi_l]] = \mathbb{E}\left[-\frac{\xi_l^2}{L-1}\right] = -\frac{1}{L-1} \tag{9}$$

1. For both estimators we have $\mathbb{E}\left[\hat{\tau}|\{W_{il}\}_{il}, \{\mu_l\}_l\right] = \frac{1}{L}\sum_l \mu_l\xi_l + \tau$. For the second design $\xi_l = 0$ and thus $\hat{\tau}$ is conditionally unbiased (and hence unconditionally as well). For the first design we can use the fact that $\mathbb{E}[\xi_l] = 0$ to get the following:

$$\mathbb{E}\left[\hat{\tau}|\{\mu_l\}_l\right] = \tau \tag{10}$$

[4] For the variance use that fact that $\left(\gamma_{il}^{(1)} - \gamma_{il}^{(0)}\right)^2 = \frac{(nL)^2}{\left(\sum_{jk}\{W_{jk}=w\}\right)^2} = 4$.

[5] In order to show this one can use that $\mathbb{E}[\sum_{k\neq l}\xi_k|\xi_l] = -\xi_l$ and $\mathbb{E}[\xi_j|\xi_l] = \mathbb{E}[\xi_k|\xi_l]$ for any $j, k \neq l$.

2. Since for the second design $\hat{\tau}$ is conditionally unbiased we have that the variance is equal to the expectation of the conditional variance:

$$\mathbb{E}\left[\mathbb{V}\left[\hat{\tau}|\{W_{il}\}_{il},\{\mu_l\}_l\right]|\{\mu_l\}_l\right]=\frac{4\sigma^2}{nL} \tag{11}$$

For the first design the variance is greater because we also have the variance of the conditional expectation:

$$\mathbb{V}\left[\frac{1}{L}\sum_l \mu_l\xi_l|\{\mu_l\}_{l=1}^L\right]=\frac{1}{L^2}\left(\sum_{l=1}^L \mu_l^2-\frac{2}{L-1}\sum_{k>j}\mu_j\mu_k\right) \tag{12}$$

and thus the total variance is equal to:

$$\mathbb{V}[\hat{\tau}|\{\mu_l\}_{l=1}^L]=\frac{4\sigma^2}{nL}+\frac{1}{L^2}\left(\sum_{l=1}^L \mu_l^2-\frac{2}{L-1}\sum_{k>j}\mu_j\mu_k\right) \tag{13}$$

Clearly the first design is less efficient than the second one.

3. Both estimators are unbiased, so they are consistent if the variance goes to zero. Clearly, this is true for the second design. For the first design one can show that the additional component is of the order $\frac{\|\mu-\overline{\mu}\|_2^2}{L^2}$ in the worst case. As a result, the answer depends on how large $\|\mu-\overline{\mu}\|_2^2$ is compared to L^2. For example, if μ_l are *i.i.d.* random variables then $\|\mu-\overline{\mu}\|_2^2$ is of order $O(L)$.

4. For the second design the variance clearly goes to zero. For the first one it never goes to zero no matter how large n is.

Задача 5[6]

Есть случайная выборка $\{(Y_i,W_i,X_i)\}_{i=1}^n$, полученная с помощью симуляций:

$$\begin{aligned}
&W_i\sim \text{Bernoulli with probability }\pi\\
&X_i|W_i\sim\mathcal{N}(0,\Sigma_X)\\
&Y_i|X_i,W_i\sim\mathcal{N}(m(X_i,W_i),\sigma^2)\\
&m(X_i,W_i)=\alpha_0+\beta_0^TX_i+\tau_0W_i+\gamma_0^TX_iW_i
\end{aligned}$$

где X_i – p-мерный вектор, а Σ_X его ковариационная матрица, W_i – бинарная переменная и Y_i – переменная интереса. Обратите внимание, что по условию W_i и X_i независимы.

Ваша цель - оценить τ. Вы рассматриваете две разных МНК оценки τ. Пусть $\hat{\tau}_{simple}$ - МНК оценка τ в регрессии:

$$Y_i=\alpha+\tau W_i+u_i$$

а $\hat{\tau}_{comp}$ - МНК оценка τ в регрессии:

$$Y_i=\alpha+\tau W_i+\beta^TX_i+\nu_i \tag{14}$$

1. Покажите, что $\hat{\tau}_{simple}$ является несмещенной оценкой τ_0. А что можно сказать о $\hat{\tau}_{comp}$?

2. Покажите, что обе оценки τ_0 являются состоятельными.

3. Пусть $\gamma_0=0$. Покажите, что $\hat{\tau}_{comp}$ более эффективна, чем $\hat{\tau}_{simple}$ (имеет меньшую асимптотическую дисперсию).

4. Пусть $\gamma_0\neq 0$. Является ли $\hat{\tau}_{comp}$ все еще более эффективной, чем $\hat{\tau}_{simple}$?

[6] По мотивам статьи “On regression adjustments in experiments with several treatments” by David A. Freedman

5. Представьте, что вы на научном семинаре и ваш одногруппник представляет эмпирические результаты, полученные с помощью спецификации (14). W_i - индикаторная переменная нахождения в группе воздействия, X_i - наблюдаемые контрольные переменные, Y_i - переменная интереса. Данные получены с помощью случайного эксперимента, то есть W_i не зависит от X_i по построению. Кто-то из аудитории комментирует: “Вы предполагаете, что γ_0 равна нулю, хотя на самом деле это не так. Поэтому ваши результаты невалидны”. Презентующий отвечает: “Данные получены с помощью случайного эксперимента, поэтому не важно, равна ли γ_0 нулю. Моя оценка верная. Используя контрольные переменные, я делаю ее более эффективной, чем простая разность средних”. Что вы думаете об этом ответе, опираясь на результаты пунктов 1-4?

Решение задачи 5

1. For the first estimator we have the following:

$$\hat{\tau}_{simple} = \tau_0 + \beta_0^T \frac{\sum_{i=1} X_i(W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} + \gamma_0^T \frac{\sum_{i=1} X_i W_i (W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} + \frac{\sum_{i=1} \varepsilon_i (W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} \tag{15}$$

Since $\{W_i\}$ is independent of ε_i and X_i and both X_i and ε_i have mean zero we get the following result:[7]

$$\mathbb{E}[\hat{\tau}_{simple} | \{W_i\}] = \tau_0 \tag{16}$$

For the OLS estimator we have the following representation:

$$\hat{\tau}_{OLS} := \tau_0 + \gamma_0^T \frac{\sum_{i=1}^n X_i W_i (W_i - \hat{W}_i)}{\sum_{i=1}^n (W_i - \hat{W}_i)^2} + \frac{\sum_{i=1}^n \varepsilon_i (W_i - \hat{W}_i)}{\sum_{i=1}^n (W_i - \hat{W}_i)^2} \tag{17}$$

where $\hat{W}_i$ is the projection of W_i on $1, X_i$. Conditional expectation has the following form:

$$\mathbb{E}[\hat{\tau} | \{W_i, X_i\}] = \tau_0 + \gamma_0^T \frac{\sum_{i=1}^n X_i W_i (W_i - \hat{W}_i)}{\sum_{i=1}^n (W_i - \hat{W}_i)^2} \tag{18}$$

In general the second part does not have (unconditional) expectation equal to zero and thus the estimator is not unbiased.

2. Using LLN and continuous mapping theorem we get the following (because X_i and ε_i have expectation zero and are independent of W_i):

$$\begin{aligned} &\frac{\sum_{i=1} X_i(W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} \to_p 0 \\ &\frac{\sum_{i=1} X_i W_i (W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} \to_p 0 \\ &\frac{\sum_{i=1} \varepsilon_i (W_i - \overline{W})}{\sum_{i=1}^n (W_i - \overline{W})^2} \to_p 0 \end{aligned} \tag{19}$$

This implies consistency.

[7] This is almost correct because W_i can be constant and then the estimator is not well-defined. We abstract away from this.

For the second estimator we have the following (follows by LLN, continuous mapping theorem and the fact that OLS projection coefficients converge to their population analogs)

$$\hat{\tau}_{OLS} := \tau_0 + \gamma_0^T \frac{\frac{1}{n}\sum_{i=1}^n X_i W_i (W_i - \tilde{W}_i)}{\pi(1-\pi)} + \frac{\frac{1}{n}\sum_{i=1}^n \varepsilon_i (W_i - \tilde{W}_i)}{\pi(1-\pi)} + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{20}$$

where $\tilde{W}_i$ are population analogs of $\hat{W}_i$. Since W_i is independent of X_i we have that $\tilde{W}_i = \pi$. Consistency then follows by the same logic as for the first estimator.

3. One can show that the difference in means has the following asymptotic representation:

$$\begin{aligned}\hat{\tau}_{simple} = \tau_0 + \beta_0^T \frac{\frac{1}{n}\sum_{i=1} X_i (W_i - \pi)}{\pi(1-\pi)} &+ \\ \gamma_0^T \frac{\frac{1}{n}\sum_{i=1} X_i W_i (W_i - \pi)}{\pi(1-\pi)} &+ \\ \frac{\frac{1}{n}\sum_{i=1} \varepsilon_i (W_i - \pi)}{\pi(1-\pi)} + o_p\left(\frac{1}{\sqrt{n}}\right) &= \\ \tau_0 + \frac{1}{\sqrt{n}}\xi_1 + \frac{1}{\sqrt{n}}\xi_2 + \frac{1}{\sqrt{n}}\xi_3 + o_p\left(\frac{1}{\sqrt{n}}\right)\end{aligned} \tag{21}$$

and the representation for $\hat{\tau}_{OLS}$ is the same, but without ξ_1 term. If γ_0 is equal to zero, then the asymptotic variance of $\hat{\tau}_{OLS}$ is equal to $\mathbb{E}[\xi_3^2]$ and the asymptotic variance of $\hat{\tau}_{simple}$ is equal to $\mathbb{E}[(\xi_1 + \xi_3)^2] = \mathbb{E}[\xi_1^2 + \xi_3^2]$ (because $\mathbb{E}[\xi_1 \xi_3] = 0$). As a result, OLS estimator is more efficient.

4. The answer is no. The difference in asymptotic variances between $\hat{\tau}_{simple}$ and $\hat{\tau}_{OLS}$ is equal to the following:

$$\beta_0^T \Sigma_X \beta_0 \frac{1}{\pi(1-\pi)} + 2\beta_0^T \Sigma_X \gamma_0 \frac{\mathbb{E}[(W_i - \pi)^2 W_i]}{\pi^2(1-\pi)^2} \tag{22}$$

If we set $\gamma_0 = -c\beta_0$ then this is equal to the following:

$$\beta_0^T \Sigma_X \beta_0 \left(\frac{1}{\pi(1-\pi)} - 2c\frac{\mathbb{E}[(W_i - \pi)^2 W_i]}{\pi^2(1-\pi)^2}\right) \tag{23}$$

For the large values of c this is negative and thus the simple estimator is better. The intuition is that $\beta_0^T X_i$ and $\gamma_0^T X_i W_i$ compensate each other on average, but when we take away the first part, the second part is not compensated anymore and thus the variance is larger.

5. The presenter is only partially correct: it is true that the estimator is always consistent, despite the fact that the linear regression is misspecified. At the same time the OLS estimator might actually be worse than the simple difference in means.