1.

a)

VARIABLES	(1) re78
treat	1,685***
	(647.6)
2.pscore_interval	-5,912
	(6,681)
3.pscore_interval	-6,846
	(6,604)
4.pscore_interval	-6,622
	(6,627)
5.pscore_interval	-5,691
	(6,635)
6.pscore_interval	-4,761
	(6,764)
Constant	11,012*
	(6,585)
Observations	445
R-squared	0.027

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To estimate the treatment effect across different intervals of the propensity score using a single regression, we can adopt an approach that integrates matching and stratification. The idea is to create a regression model that accounts for varying treatment effects over distinct propensity score intervals. By doing so, we can observe how the treatment effect shifts as the propensity score changes, giving a more nuanced understanding of the causal effect.

First, we need to stratify the sample into intervals of the propensity score, such as 0.1-width bins (e.g., 0.0–0.1, 0.1–0.2, etc.). Each interval is represented by an indicator variable that flags which

observations fall within a specific range of p(X). This stratification allows the model to isolate and estimate effects within different sections of the propensity score distribution.

We include interaction terms between the treatment variable and these indicator variables. This would allow the effect of the treatment to vary across different levels of p(X).

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j I_j + \gamma T + \sum_{j=1}^{k} \delta_j (T \times I_j) + \epsilon$$

I_j are the dummy indicators for each propensity score interval.

T is the treatment variable.

δ_j represents the interaction terms that show how the treatment effect varies across each propensity score interval.

The primary parameters of interest would be the δ_j coefficients. Each δ_j provides the estimated treatment effect for the corresponding propensity score interval. By examining these coefficients, we could observe how the treatment effect changes as the propensity score varies.

b)

The two main assumptions we need are:

Conditional Independence Assumption (CIA): This assumption implies that, conditional on the set of observed covariates, receiving the job training (treatment) is independent of the potential outcomes (earnings). This means that, after controlling for these covariates, there are no unobserved factors that affect both the likelihood of receiving training and the income outcome.

Common Support (Overlap): For the causal inference to be valid, there must be a substantial overlap in the distribution of propensity scores between the treatment and control groups. This ensures that for each treated individual, there is a similar non-treated individual with a comparable propensity score.

c)

The method of dividing propensity scores into intervals of 0.1 and estimating regressions within each interval helps address the inflexibility concern by allowing for localized estimates of the treatment effect. This means that the model can capture heterogeneity in treatment effects across different

sections of the propensity score distribution, making the analysis more responsive to variations in how treatment effects might differ across subgroups.

By including dummy variables for each propensity score interval (i.pscore_interval), the approach can flexibly adjust for differences across these intervals. This step incorporates a form of interaction between the treatment variable (treat) and different levels of the propensity score, thus allowing the estimation to reflect potential variation in treatment effects based on the estimated likelihood of treatment.

Estimating the treatment effect within propensity score intervals effectively acts as a piecewise model. Each interval's regression can adapt to the local relationship between treatment and outcome, which captures non-linearities that a single, unified model might miss. This piecewise approach can be seen as an intermediate step between a simple matching approach and a more flexible, non-parametric estimation.

The main critique of standard propensity score matching methods—rigidity in post-matching effect estimation—is addressed here by allowing for heterogeneity in treatment effects across multiple subgroups. While this method is more flexible than a simple average treatment effect estimation post-matching, it still retains some limitations in the assumption of linear relationships within each block. However, it is a practical step towards more flexible modeling while maintaining interpretability.

d)

Matching methods rely on the assumption that, after matching on observed covariates, the treatment assignment is as good as random (CIA). This assumption is crucial for the matching process to remove confounding and mimic a randomized controlled trial (RCT). Without this assumption, matching may not account for unobserved confounders that influence both treatment and outcomes, leading to biased estimates.

Directed Acyclic Graphs (DAGs) help identify and visualize the relationships between variables, including potential confounders and mediators. Understanding these pathways ensures that researchers correctly adjust for variables. For instance, matching on a mediator rather than a true confounder can distort the estimated treatment effect. DAGs clarify these distinctions and ensure researchers adjust for the right covariates.

Matching only balances observed covariates. If there are unobserved confounders (variables that influence both the treatment and outcome but are not accounted for), the estimated treatment effect will still be biased. This highlights the need for assumptions such as “no hidden confounding,” which cannot be verified by the matching method alone.

Example:

Evaluating the impact of a job training program on future earnings. If an unmeasured variable, such as inherent motivation, influences both the likelihood of attending job training and the future earnings, matching on observed variables alone will not yield a true causal effect. Here, the results would be biased because the matching method failed to account for an unobserved confounder. If individuals self-select into job training based on characteristics linked to better job prospects (e.g., individuals who are already more motivated or have better social networks), matching on observed factors like age and prior income won't eliminate this bias. The causal assumptions would help identify this potential bias, guiding adjustments or sensitivity analyses.

A DAG can show paths between treatment and outcome, such as backdoor paths that need to be closed (e.g., controlling for common causes) or colliders that must not be conditioned on to prevent bias. Without causal assumptions, one might mistakenly match on variables that introduce bias (e.g., matching on mediators rather than confounders).

e)

According to my results 183 out of 185 treatment units get matched with at least one unit in the control group. And 260 out of 260 control units get matched.

The tables are available in excel files provided with the homework, made using the stata codes.

f)

VARIABLES	(1) re78
treat	1,648**
	(692.3)
age	66.18
	(45.53)
education	430.9**
	(178.3)
re74	0.115
	(0.0847)
re75	-0.0342
	(0.142)
Constant	-1,636
	(2,273)
Observations	366
R-squared	0.042

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The estimated effect of the variable "treat" on "re78" is 1,648, and it is statistically significant at the 5% level, as indicated by the two asterisks (**). The standard error associated with this estimate is 692.3.

g)

According to my results 131 out of 185 treatment units get matched with at least one unit in the control group. And 196 out of 260 control units get matched.

The tables are available in excel files provided with the homework, made using the stata codes.

h)

We get a summery of re74 in both of the cases to get a better glance of what we're working with:

The Matched Data

Variable	Obs	Mean	Std. Dev.	Min	Max
re74 treat/control	131	20.655	172.519	0	1716.509

The Real Data (Unmatched)

Variable	Obs	Mean	Std. Dev.	Min	Max
re74	445	2102.265	5363.582	0	39570.68

As we can see from the summary tables above, one notable observation is that many of the records in the matched data show zero real earnings in 1974. This likely indicates that the treatment group of the job training program did not have any earnings before the program began, suggesting that the program targets individuals who are entering the workforce rather than those already employed.

This observation aligns with what the paper states about the job training program. The program admitted women receiving aid for families with dependent children, recovering addicts, released offenders, and men and women who had not completed high school—all groups typically without prior earnings.

i)

Exact matching requires precise matches between treated and control units across all covariates. This can limit the number of observations that can be included in the analysis as we see from our analysis, reducing the generalizability of the findings. In cases where exact matches are scarce, a significant portion of the data may be discarded, leading to a potential loss of valuable information. As the number of covariates increases (i.e., in high-dimensional data), finding exact matches becomes exponentially more difficult. This issue, often called the "curse of dimensionality," can severely constrain the practical application of exact matching. If exact matches are not feasible for some treated units, it might introduce bias due to non-representative samples. This situation could lead to estimates that do not fully reflect the true causal effect.

j)

Nearest Neighbor Matching:

```
Treatment-effects estimation                   Number of obs      =        445
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Distance metric: Mahalanobis                               max =        154
```

re78	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
treat (1 vs 0)	1841.693	682.3879	2.70	0.007	504.237	3179.148

	Raw	Matched
Number of obs =	445	370
Treated obs =	185	185
Control obs =	260	185

	Standardized differences		Variance ratio	
	Raw	Matched	Raw	Matched
black	.0438866	3.05e-16	.9250286	1
hispanic	-.1745611	-8.78e-17	.5828804	1
married	.0936407	-7.07e-17	1.180212	1
nodegree	-.3039864	-.0238484	1.499755	1.022846

Caliper Matching

```
Treatment-effects estimation                   Number of obs      =        445
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                                  min =          1
Treatment model: logit                                     max =        154
```

re78	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
treat (1 vs 0)	1792.922	685.2617	2.62	0.009	449.8332	3136.01

	Raw	Matched
Number of obs =	445	370
Treated obs =	185	185
Control obs =	260	185

	Standardized differences		Variance ratio	
	Raw	Matched	Raw	Matched
black	.0438866	3.05e-16	.9250286	1
hispanic	-.1745611	-8.78e-17	.5828804	1
married	.0936407	.0278399	1.180212	1.046651
nodegree	-.3039864	4.87e-16	1.499755	1

Choosing the best matching method in econometrics requires careful evaluation of several key aspects. First and foremost, the choice of method should align with the underlying assumptions of causal

inference. When a study aims to estimate the causal effect, it must ensure minimal selection bias or use methods designed to correct for such biases. This alignment helps in obtaining valid and reliable causal estimates.

Balancing between treated and untreated groups is another critical consideration. Matching methods are designed to create such balance, thereby mimicking the conditions of a randomized experiment. This is essential for estimating treatment effects as it reduces potential biases that arise from non-random treatment assignment. A method that effectively balances these groups is necessary to achieve comparable samples.

The type of data available also plays a significant role in selecting an appropriate matching method. When there are significant differences in the covariate distributions between treated and untreated groups, methods such as propensity score matching or kernel matching can be more suitable than simple matching. These methods adjust for covariate imbalances, enhancing the comparability of treatment groups.

i)

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	6349.1435	4554.80112	1794.34238	632.853392	2.84
	ATT	6349.1435	2447.74148	3901.40203	1721.44463	2.27

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	260	260
Treated	185	185
Total	445	445

The Mahalanobis matching method is a non-parametric approach used for causal inference to estimate the effect of a treatment by matching units based on the similarity of their covariates. One major advantage of this method is its ability to handle multidimensional matching effectively. By considering the correlation between variables, Mahalanobis matching offers an improvement over simple distance metrics like Euclidean distance, allowing for better matching in high-dimensional spaces. This method can also help achieve better balance between treated and control groups by accounting for the covariance structure among covariates, which can reduce bias in treatment effect estimation. Additionally, Mahalanobis matching does not require assuming any specific functional form between covariates and outcomes, offering more flexibility and robustness in certain applications.

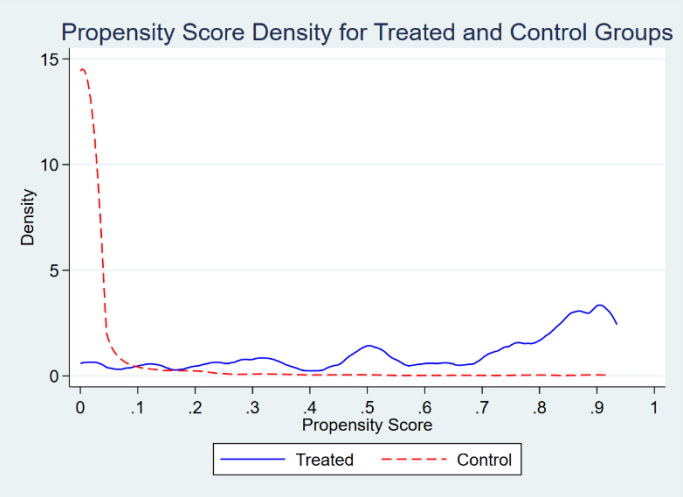
However, there are some limitations to consider. Mahalanobis matching can become computationally intensive, especially as the number of covariates or dataset size increases, which can make it less feasible for large-scale studies. Another challenge is the "curse of dimensionality," where the performance of the matching deteriorates as the dimensionality of the covariate space increases, making it difficult to find sufficiently similar matches. The method's effectiveness also depends on the distribution of the covariates; if they have skewed or non-standard distributions, Mahalanobis distance may not accurately reflect true similarity, leading to suboptimal matches. Lastly, the method does not inherently ensure common support between treated and control groups, so finding appropriate matches can be challenging, particularly with limited sample sizes.

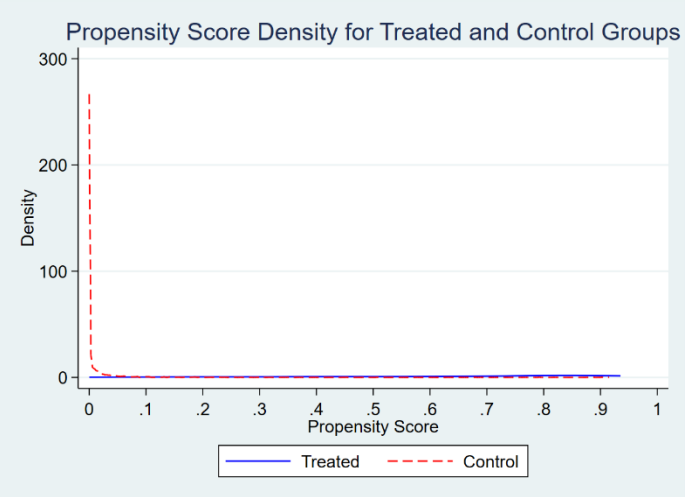
Despite its advantages in handling multidimensional data without strong assumptions, researchers need to be mindful of these computational and practical limitations when using Mahalanobis matching.

2.

a)

- Average Propensity Score for Treated: **0.63639427**
- Average Propensity Score for Control: **0.02701488**





The table shows that the average propensity score for the treated group is significantly higher than that of the control group, indicating a difference in treatment assignment probabilities between the two groups.

b)

Metric	Weighted Mean Outcome for Treat	Weighted Mean Outcome for Control	ATE
IPW	7677.685	20793.28	-13115.59
SIPW	7804.355	19534.98	-11730.62

IPW estimate for treated mean: 7677.685
IPW estimate for control mean: 20793.28
ATE (IPW): IPW-Treatment Minus IPW-Control

SIPW estimate for treated mean: 7804.355
SIPW estimate for control mean: 19534.98
ATE (SIPW): SIPW-Treatment Minus SIPW-Control

c)

VARIABLES	(1) re78
treat	751.9
	(915.3)
age	-83.57***
	(20.81)
education	592.6***
	(103.3)
hispanic	2,163**
	(1,092)
black	-570.9
	(495.2)
married	1,241**
	(586.3)
nodegree	590.5
	(646.8)
re74	0.278***
	(0.0279)
re75	0.568***
	(0.0276)
Constant	-129.7
	(1,689)
Observations	2,675
R-squared	0.586

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As is shown in the table above the ATE in this situation equals 751.9 which has a positive sign and is much smaller than the negative effect estimated in the previous part of the question. The difference in treatment effect estimates between the linear regression model and the IPW\SIPW model arises due to the underlying assumptions and methods each approach employs. Linear regression assumes a specific functional form for the relationship between the outcome and predictors, typically a linear relationship. This method adjusts for covariates directly in the model, implying that the treatment effect is conditional on these covariates. In contrast, IPW does not assume a specific model for the outcome. Instead, it assigns weights to the observations to create a pseudo-population where treatment assignment is independent of observed covariates, allowing for a more direct estimation of the average treatment effect (ATE).

One key reason for the difference is how each method handles the balancing of covariates. IPW seeks to balance covariates between treated and control groups by applying weights based on the propensity score, reducing potential bias due to baseline differences. This approach simulates a randomized trial environment if the propensity score model is correctly specified. On the other hand, linear regression controls for covariates by estimating the conditional mean of the outcome. If the model is misspecified or the relationship between covariates and the outcome is non-linear, the treatment effect estimate may be biased.

The estimation of the treatment effect itself also varies between the two methods. IPW directly estimates the ATE by weighting individuals according to their propensity scores, offering robustness against certain types of model misspecification. In contrast, linear regression provides an estimate conditional on the included covariates and may suffer if the treatment assignment mechanism is correlated with unobserved factors affecting the outcome. Additionally, if there is multicollinearity among predictors, the estimate from regression may differ from the true ATE.

Finally, unobserved heterogeneity plays a role in the difference between the two approaches. IPW may yield a more accurate ATE if the propensity score model accurately captures the treatment assignment process. In contrast, linear regression could produce biased estimates if there are unobserved confounders impacting both treatment and outcome, as it can only adjust for observed variables. These distinctions highlight why the treatment effect estimates from IPW and linear regression can differ in practice.

d)

Metric	Weighted Mean Outcome for Treat	Weighted Mean Outcome for Control	ATE
IPW	6596.069	6300.237	295.8316
SIPW	6266.615	5560.766	705.849

IPW estimate for treated mean: 6596.069
IPW estimate for control mean: 6300.237
ATE (IPW): IPW-Treatment Minus IPW-Control

SIPW estimate for treated mean: 6266.615
SIPW estimate for control mean: 5560.766
ATE (SIPW): SIPW-Treatment Minus SIPW-Control
As we can see from our results when we trim the p-score we got almost the same treatment effect of the linear regression model from IPW and a closer than before estimate on the treatment effect from SIPW both of which now have the same sign with our linear model estimate.

e)

For Blacks:

IPW estimate for treated mean: 6344.594
IPW estimate for control mean: 6477.756
SIPW estimate for treated mean: 6009.174
SIPW estimate for control mean: 5614.634

Metric	Weighted Mean Outcome for Treat	Weighted Mean Outcome for Control	ATE
IPW	6344.594	6477.756	-133.162
SIPW	6009.174	5614.634	394.5392

For non-Blacks:

IPW estimate for treated mean: 6936.651
IPW estimate for control mean: 5960.731
SIPW estimate for treated mean: 6876.92
SIPW estimate for control mean: 5351.519

Metric	Weighted Mean Outcome for Treat	Weighted Mean Outcome for Control	ATE
IPW	6936.651	5960.731	975.9195
SIPW	6876.92	5351.519	1525.401

From our observations we can see that both IPW and SIPW ATEs for Blacks are significantly lower compared to the full sample. The IPW ATE is even negative, indicating a potential adverse treatment effect or different dynamics in the treated versus control groups within this subgroup. Both ATEs for non-Blacks are higher than the full sample's ATEs, suggesting that the treatment effect is more pronounced in this subgroup compared to the general population.

f)

Outcome Variable	(1) I (Diff-re78-re75)	(2) II (re78)
treat	2,326.5049***	-581.8331
	(813.8592)	(841.2555)
re75		0.8341***
		(0.0154)
Constant	2,490.5833***	5,653.0804***
	(214.0293)	(360.4367)
Observations	2,675	2,675
R-squared	0.0030	0.5529

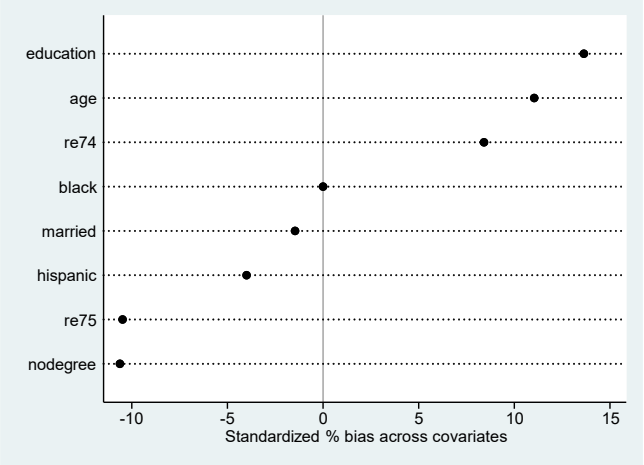
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

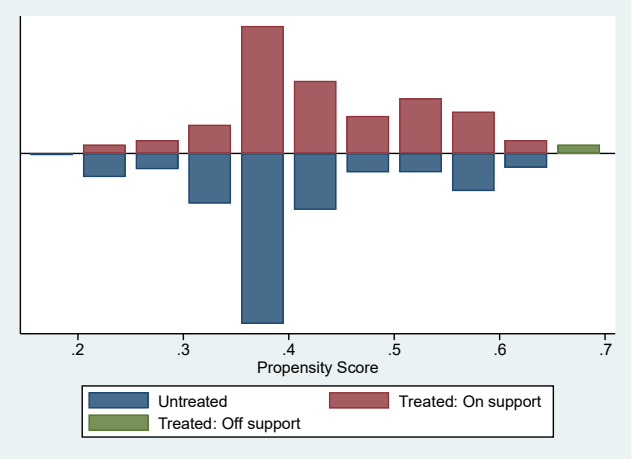
The first model shows a significant positive treatment effect on the difference in earnings between 1978 and 1975. However, in the second model, after controlling for prior earnings in 1975, the treatment effect on 1978 earnings is not statistically significant.

We could use other specifications to better our estimates like including interaction terms, using other control covariates, using the log of earning and etc.

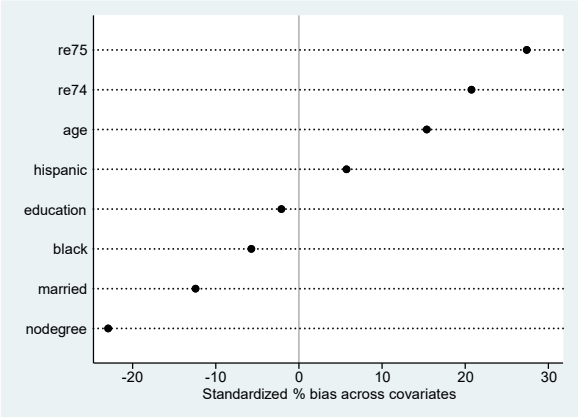
3.

a) NSW

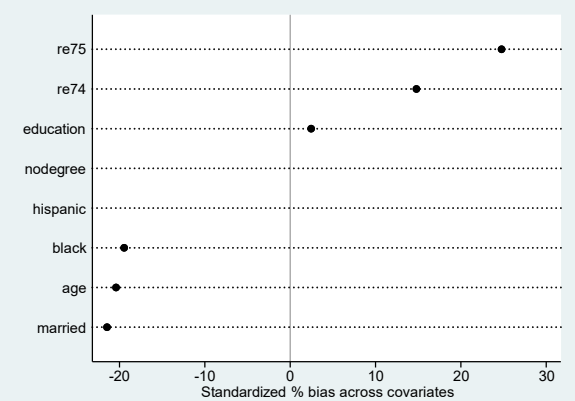




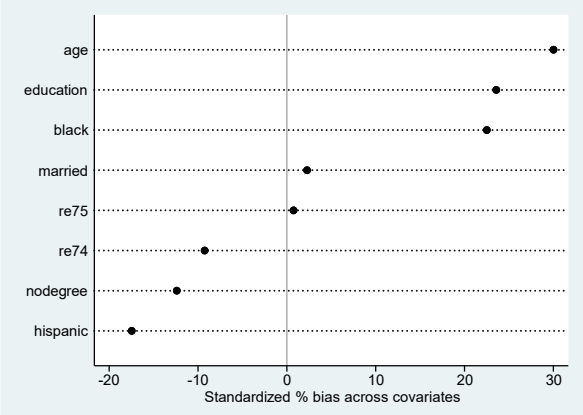
First Block:



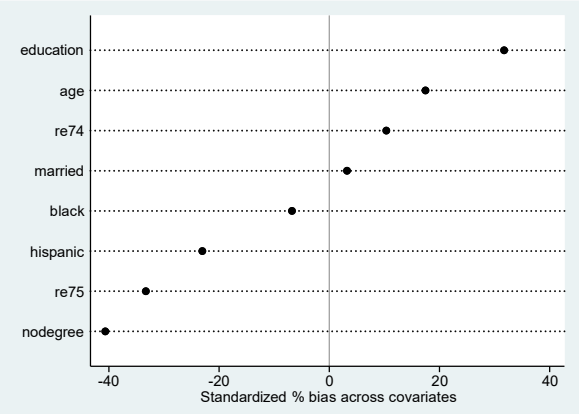
Second Block:



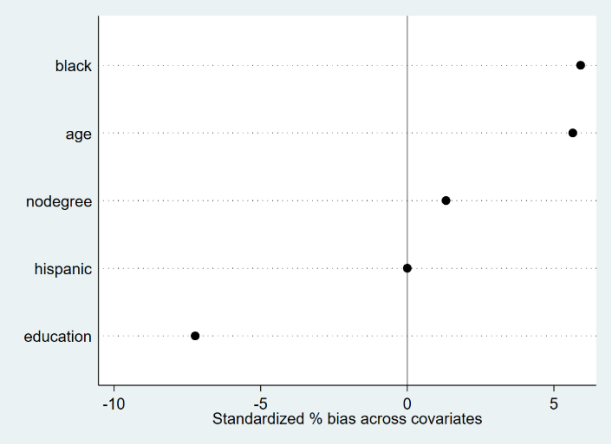
Third Block:

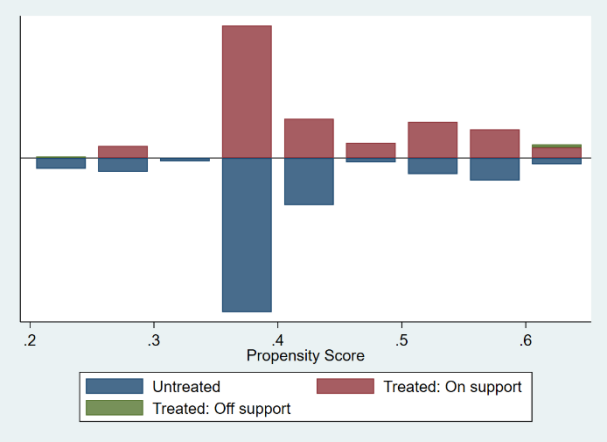


Forth Block:

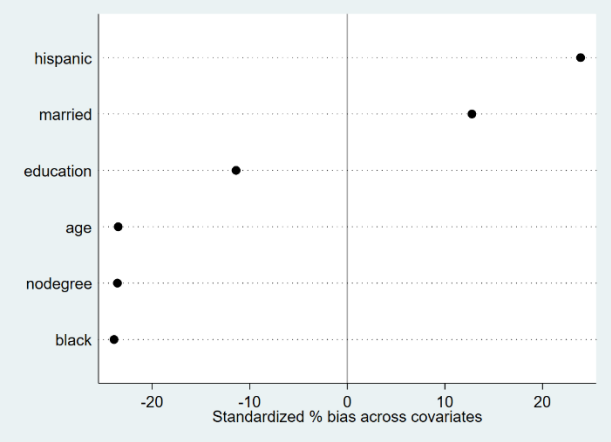


After removing re74 re75:

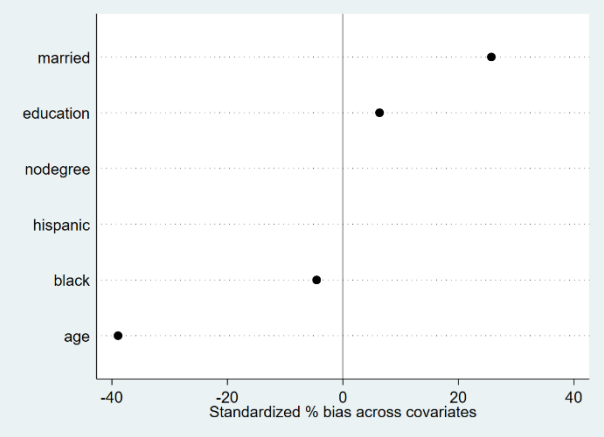




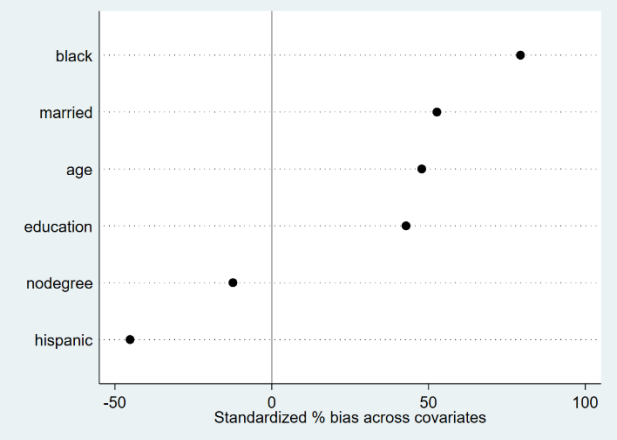
First Block:



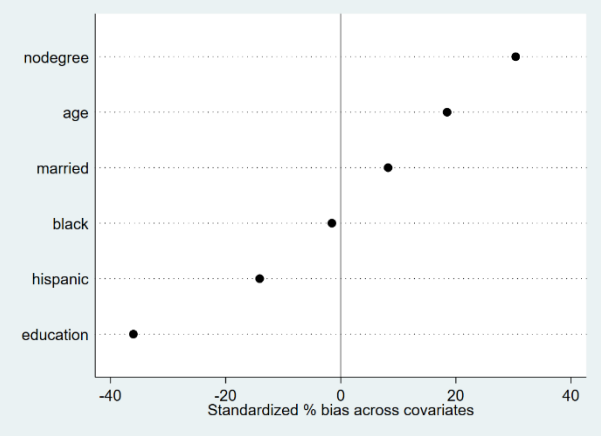
Second Block:



Third Block:



Forth Block:



As we see in the graphs when we remove re74 and re75 from our analysis the balancing property gets better.

Dropping variables to satisfy the balancing property may occur when those variables contribute to imbalance in the distribution of covariates across treatment and control groups. This action is taken to achieve balance, which is crucial for ensuring that comparisons between treated and untreated groups are unbiased and reflect the causal effect of treatment.

This implies that the interpretation of propensity scores should be approached with caution. Specifically, removing variables might mean that the propensity score model only partially captures the selection process into treatment. As a result, the propensity score may no longer represent the full range of factors influencing treatment assignment. The scores will still be useful for balancing observed covariates, but they may overlook unmeasured variables or effects not included in the model.

b)

10 quantiles of pscore	treat 0	1	Total
1	35	10	45
2	27	17	44
3	31	16	47
4	28	14	42
5	30	15	45
6	26	18	44
7	22	23	45
8	20	24	44
9	17	29	46
10	24	19	43
Total	260	185	445

The table presents the distribution of observations between control (treat = 0) and treatment (treat = 1) groups across 10 quantiles of propensity scores. Each quantile contains a relatively equal number of total observations, with a breakdown provided for both groups. This distribution helps us assess how well the treatment and control groups overlap in terms of their propensity scores, which is essential for matching and ensuring reliable causal inference.

In the upper half of the distribution (quantiles 1 to 5), there is a noticeable imbalance between the control and treatment groups. For instance, the first quantile shows 35 observations in the control group but only 10 in the treatment group. This trend continues in the subsequent quantiles, where the number of control observations consistently outnumbers those in the treatment group. Such an imbalance suggests that for these lower propensity score deciles, there is limited overlap between treated and untreated units. This lack of overlap can be problematic, as it may lead to insufficient matching and potentially biased estimates of treatment effects for those with lower scores.

In contrast, the lower half of the table (quantiles 6 to 10) shows a more balanced distribution between the two groups. For example, quantile 7 has 22 control and 23 treatment observations, demonstrating a near-equal representation. Similar patterns are seen in other higher quantiles, such as quantile 9, which has 17 control and 29 treatment observations. This improved balance indicates better overlap between the treatment and control groups in the higher range of propensity scores, suggesting that matching is likely more feasible and robust in these deciles.

The overall takeaway from this distribution is that while there is some degree of balance in the higher propensity score deciles, the lower deciles exhibit a significant mismatch between the treatment and

control groups. This imbalance may require further attention. To address this, you might consider trimming the sample or focusing the analysis on propensity score ranges where there is better overlap. Alternatively, more advanced matching techniques, such as caliper or radius matching, could be employed to ensure that treated and control units are more closely aligned in terms of their propensity scores.

c)

treat	Mean	SD	Min	Max
0	.4000215	.0936532	.1716353	.6392755
1	.4378076	.0949339	.231849	.6756708
Total	.4157303	.0959106	.1716353	.6756708

The OLS regression to estimate the probability of treatment offers an insightful preliminary analysis by illustrating how covariates influence the likelihood of being assigned to the treatment group. In this dataset, the comparison between control and treatment groups shows that the predicted probabilities differ slightly between the two. Specifically, for the control group (where treat = 0), the average predicted probability of treatment is approximately 0.40. This indicates that individuals in the control group have, on average, a 40% estimated probability of being in the treatment group. On the other hand, the treatment group (where treat = 1) has an average predicted probability of about 0.44, suggesting that individuals in the treatment group have a slightly higher estimated probability of treatment compared to the control group.

The use of OLS for this type of analysis is beneficial due to its simplicity and ease of interpretation. The coefficients obtained from an OLS regression provide a direct indication of how each predictor affects the probability of treatment. This makes it possible to quickly identify which factors are associated with an increased or decreased likelihood of being in the treatment group. Additionally, OLS allows for a linear approximation that can be useful for understanding relationships in the dataset at an initial stage.

However, there are inherent limitations when using OLS to estimate probabilities. One significant issue is that OLS does not constrain the predicted probabilities to the range of 0 to 1, which are the logical bounds for probabilities. Although in this particular dataset, the predicted values stayed within this range, there is no guarantee that this will always be the case. Predicted probabilities outside of 0 and 1 would be nonsensical and problematic for interpretation. Another limitation is that OLS assumes a linear relationship between the covariates and the outcome, which may not capture more complex,

non-linear relationships in the data. This assumption can lead to inaccuracies if the true relationship is not linear.

Overall, while OLS provides a straightforward and interpretable way to assess the influence of covariates on treatment probability, it is generally more appropriate to use logistic regression for such analyses. Logistic regression models the log-odds of treatment as a function of the covariates, ensuring that predicted probabilities remain within valid bounds and better capturing potential non-linear relationships.

d)

VARIABLES	(1) re78
o.pscore_0	-
pscore__1	6,533 (6,607)
pscore__2	526.9 (1,387)
treat	1,832*** (658.9)
o.pscore_treat_0	-
o.pscore_treat__1	-
pscore_treat__2	182.3 (2,553)
Constant	4,479*** (431.0)
Observations	445
R-squared	0.021

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As can be seen from the table above the treatment coefficient is significant and has a very low p-value.

e)

VARIABLES	(1) re78
o.pscore_0	-
pscore__1	5,056
	(7,274)
pscore__2	-2,873
	(2,315)
treat	1,689**
	(662.0)
o.pscore_treat_0	-
o.pscore_treat__1	-
pscore_treat__2	452.6
	(2,533)
age	55.53
	(45.32)
education	445.3*
	(229.7)
black	-2,185*
	(1,170)
hispanic	2,277
	(2,283)
married	-146.6
	(880.3)
nodegree	231.7
	(1,035)
re74	0.105
	(0.0948)
re75	0.0126
	(0.154)
Constant	82.13
	(3,410)
Observations	445
R-squared	0.060

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As can be seen from the table above the treatment coefficient is still significant but not as significant as the treatment effect of the last part.