Sample Final Report Evaluation

Project Title: Examining the Impact of Paid Family Leave Policy on Firm's Risk-Taking

Student Name: XXX || **Student ID #:** XXX

I. Research Question & Objectives (30% weight)

- **Clarity and Specificity of Research Question(s):**
  - *Score: 5/5*
  - **Comments:** The research question "How do state-level Paid Family Leave (PFL) policies affect firm-level risk-taking behavior?" is clear, focused, and directly addresses a relevant gap in the literature with a focus on firm behavior/response. It is highly appropriate for the scope and has potential to provide societal (beyond managerial) implications.
- **Feasibility and Significance of Objectives:**
  - *Score: 5/5*
  - **Comments:** Objectives are clear implicitly through the methodological choices (DiD, Callaway & Sant'Anna, DML, heterogeneous effects). They are achievable given the articulated data sources (e.g., Compustat) and methods. The potential for contribution is significant, exploring a less-examined area of PFL impact.

II. Methodology (35% weight)

- **Appropriateness and Justification:**
  - *Score: 4/5*
  - **Comments:**
    - The choice of Difference-in-Differences (DiD) as the core design is appropriate given the availability of the panel data and the exogenous variations in it. Your rationale for combining baseline DiD (TWFE), Callaway & Sant'Anna gt-ATT estimator, and DML along with matching estimators to bolster robustness and address the complications of staggered implementation and confounding is strong.
    - Your awareness of "quasi-experimental variation" versus a true experiment is well-founded, but the point you made during your presentation about treatment-timing variation enabling policy evaluation needs clarification. Staggered enactment of PFL is not what allows us to infer causal effects (it should rather be thought of as an empirical challenge); recall that the classic DiD setup, e.g., Card and Krueger's minimum-wage study, used a treatment without timing variation.
- **Rigour and Ethical Considerations:**
  - *Score: 5/5*

- **Comments:**
  - Detailed description of data sources (e.g., Compustat, BEA) and variable construction are articulated. The explanation of the DiD, TWFE, Callaway & Sant'Anna, and DML methods is clear and demonstrates a solid understanding of the approaches used. The use of robust standard errors is noted.
  - The inclusion of various estimators (Lasso, Ridge, RF) in DML is appropriate and aligns with the methodological guidelines standard in the literature (see Ahren et al. 2025 for a Practical Guide to Double ML); although it would help to present the model specifications of all of the methods being used (e.g,. for DML, write out the treated CEF and the outcome CEF along with the residualized regression used to estimate the confounder-adjusted ATT; for matching, write out the treatment regression used to estimate the treatment probabilities/propensity scores). Also, clarify you used PSM.
  - Ethical Considerations are not explicitly discussed (which is okay as it is typical for research using publicly available, anonymized firm-level data).

III. Results/Findings (15% weight)

- **Clarity and Presentation:**
  - *Score: 4/5*
  - **Comments:** Results are presented clearly using tables (Report pp.7-13) and summarized effectively in the slides. The progression from TWFE, Callaway & Sant'Anna, and DML is logical. The report presents a figure depicting the parallel trends assumption and the dynamic effects.
- **Interpretation and Analysis:**
  - *Score: 4/5*
  - **Comments:**
    - Your interpretation is generally sound, especially in identifying the discrepancy between the baseline DiD and Callaway results. You correctly note the insignificance of effects after incorporating fixed effects or using Callaway & Sant'Anna gt-ATT estimator.
    - While you accurately highlighted the benefits of Callaway & Sant'Anna estimator over the conventional TWFE estimator when we have staggered treatment adoption, you could deepen your discussion on the issues raised in recent econometric literature, e.g., Goodman-Bacon 2021 2023; TWFE is a weighted average of the gt ATT with sometimes *negative weights*.
      - Goodman-Bacon, Andrew, 2021. "Difference-in-differences with variation in treatment timing," Journal of Econometrics, Elsevier, vol. 225(2), pages 254-277.
    - We currently see a null finding on the PFL. Consider delving deeper into specific industry-level effects, and substantially extending your heterogeneous effects analysis to uncover where the effects *do* manifest.

This deeper exploration would provide richer insights and more actionable findings for your research and policy discussions.

IV. Discussion & Conclusion (15% weight)

- **Discussion of Findings in Relation to Literature:**
  - *Score: 4/5*
  - **Comments:** The discussion highlights the mixed results across methods and the potential reasons for discrepancies (e.g., bias in baseline DiD). It connects the findings back to the research question and briefly discusses policy implications. It also identifies heterogeneous effects by firm characteristics (SME, international exposure), which is interesting and needs further investigation.
- **Conclusion, Contributions, and Future Work:**
  - *Score: 4/5*
  - **Comments:** The conclusion provides clear key takeaways, reiterating that PFL policies can influence firm behavior, but effects are not uniform and depend on firm characteristics. It articulates the contribution of using multiple methods. "Next Steps" are well-articulated, including improving diagnostics and further refining methods, showing a good understanding of areas for improvement in a research project.

V. Writing Quality & Presentation (5% weight)

- **Organization and Structure:**
  - *Score: 5/5*
  - **Comments:** Both the report and slides are logically structured with clear headings and a coherent flow from introduction to methodology, results, and conclusion.
- **Clarity, Grammar, and Style:**
  - *Score: 3/5*
  - **Comments:** The writing in the report is generally clear, concise, and uses appropriate academic language. However, there are some grammatical errors and typos (e.g., the title of the paper spells “Frim’s” instead of “Firm’s”). These issues could detract readers and undermine the professionalism and academic rigor of the research presented, and thus should be thoroughly proofread and corrected.

General Comments:

XXX, this is a very strong and promising research project with significant potential for your thesis. Remember that you are tackling a highly relevant policy question using modern causal inference methods, which is commendable for a class project. I’ve thoroughly enjoyed reading your report and listening to your presentation.

Areas for Further Development:

1. **Interpreting Insignificant Results:** You observe that the TWFE and Callaway & Sant'Anna models show insignificant effects. Deepen your discussion on *why* this may be the case and, more importantly, continue investigating to pinpoint interesting results that we could build on (e.g., industry-specific effects).
2. **Heterogeneous Treatment Effect:** Prioritize further exploring the mechanisms behind the heterogeneous effects you observed. Why do SMEs or domestic firms respond differently?
3. **DML Diagnostics:** Address the DML diagnostics (cross-fitted R^2 and CVC) to validate your choice of learners and the robustness of those findings. Note that you should remember to use regressors for continuous nuisances (e.g. risk-taking) and classifiers for binary nuisances (e.g. treated state indicator) to ensure valid predictions and residuals.
4. **Additional Robustness Checks:** Consider also presenting results from placebo tests and sensitivity analyses to alternative control groups once you've solidified the context.

This project is on an excellent track to become a strong thesis! Keep up the great work.

Overall Score: 89.5 / 100 (This will contribute **XX / XX** to your final letter grade)
