Exit, Voice, and Response on Digital Platforms: An Empirical Investigation of Online Management Response Strategies

Online Appendix A: Analysis of Bidirectional Dynamics between Online Review Ratings and Business Owners' Responses

We employ a method of panel vector autoregression (PVAR) using restaurant level data across time in Yelp to examine the dynamics between online review ratings and business owners' responses. In the prior literature, vector autoregression (VAR) is widely used in Macroeconomics (Love and Zicchino 2006), and recently it has been increasing adopted in the Information Systems (IS) area (Chen et al. 2015). The main advantages of our PVAR approach is that (i) it allows us to examine the bi-directional relationships between online review ratings and business owners' responses through Granger causality tests; and (ii) the dynamics between online review ratings and business owners' responses over time can be assessed and visualized through techniques such as the impulse response functions.

The specification of our model has the following form:

$$Y_{i,t} = A(L)Y_{i,t} + c_i + month_t + \varepsilon_{i,t},$$

where $Y_{i,t}$ is a vector of covariates as follows:

$$Y_{i,t} = \begin{bmatrix} OwnerResponse_{it} \\ AveReviewRating_{it} \end{bmatrix},$$

c_i is the restaurant fix effect, and $month_t$ is a set of monthly time dummies. The lag operator L is defined by $LY_{it} = Y_{i,t-1}$, and we also define the symbol $L^p Y_{it} = Y_{i,t-p}$. Let $A(L)$ be the lag polynomial:

$$A(L) = a_1 L + \cdots + a_p L^p,$$

which is defined as an operator such that

$$A(L)Y_{i,t} = a_1 \cdot Y_{i,t-1} + \cdots + a_p \cdot Y_{i,t-p}.$$

In order to determine the order of lag, p, we look at different model selection measures as follows:

Table A.1 Panel VAR Model Selection

lag	MBIC	MAIC	MQIC
1	-47.01931	62.52284	28.69459
2	34.30945	107.3375	84.78538
3	66.60038	103.1144	91.83835

Based on the three model selection criteria by Andrews and Lu (2001), first-order panel VAR (p = 1) is the preferred model, since this has the smallest Modified Bayesian Information Criterion (MBIC), Modified Akaike's Information Criterion (MAIC) and Modified Quasilikelihood under the Independence Model Criterion (MQIC). Then, we estimate a first-order panel VAR model using generalized method of moments (GMM) as follows.

Table A.2 Panel VAR Estimation Results

INDEPENDENT VARIABLES	Equation (1) Dependent Variable $owner\_res\_count_{\mathrm{i,t}}$	Equation (2) Dependent Variable $av\_review\_rating_{i,t}$
$OwnerResponse_{i,t-1}$	-0.0847***	0.138***
	[0.0109]	[0.00942]
$AveReviewRating_{i,t-1}$	-0.900***	1.048***
	[0.403]	[0.0442]

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

In the following table, we look at the Granger causality test, and it shows that $OwnerResponse$ Granger-causes $AveReviewRating$, and $AveReviewRating$ Granger-causes $OwnerResponse$ at the usual confidence levels.

Table A.3 Panel VAR-Granger Causality Wald Test

Ho: Excluded variable does not Granger-cause Equation variable

Ha: Excluded variable Granger-causes Equation variable

Equation	Excluded	Chi2	df	Prob > chi
$OwnerResponse_{it}$	$AveReviewRating_{it}$	213.939	1	0.000
$AveReviewRating_{it}$	$OwnerResponse_{it}$	498.037	1	0.000

We also check the stability condition of the estimated panel VAR. The resulting table of eigenvalues confirms that the estimate is stable: All the eigenvalues lie inside the unit circle, and panel VAR satisfies stability condition.

Table A.4 Eigenvalue Stability Condition

Eigenvalue real	Eigenvalue Imaginary	Modulus
.7247558	0	.7247558
.0383453	0	.0383453

We are interested in the impact of exogenous changes in each endogenous variable to other variables in the panel VAR system. In the following figure, we estimate the impulse-response functions (IRF). The IRF confidence intervals are computed using 200 Monte Carlo draws based on the estimated panel VAR model. From the IRF plot Figure A.1b, we can see that a positive shock on the number of business owners' responses has a positive impact on Yelp rating. As time goes by, the positive impact first goes up and then decreases gradually. The IRF plot Figure A.1c shows that a current positive shock on Yelp rating has a persistent negative impact on the number of business owners' responses.

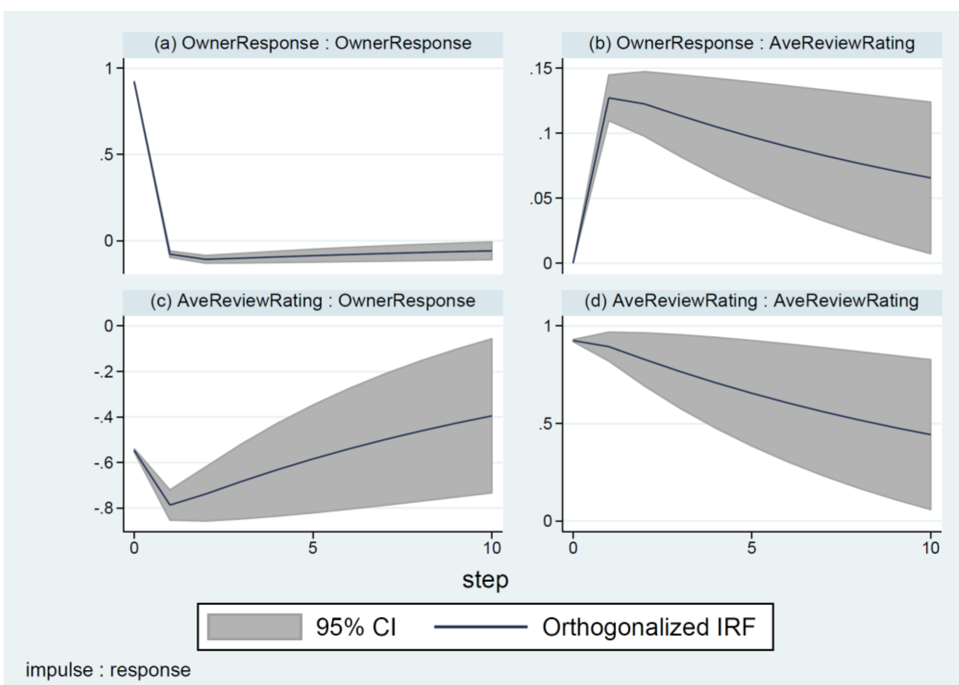

Figure A.1 Impulse-Response Functions in the PVAR Model

Online Appendix B: Quantile Regression and Subsample Analyses on Online Management Responses

A striking pattern in the data is that the number of restaurant check-ins is remarkably skewed. Quantile regression analysis is particularly useful when the conditional distribution of check-ins is not symmetric and does not have a "standard" shape (Koenker and Hallock 2001). The quantile regression models allow us to account for unobserved heterogeneity and heterogeneous covariates effects. We estimate regression equation (2) using quantile regressions. In Figure B.1, we plot the parameter estimates β_2 of the quantile regressions based on equation (2). There are four estimated quantile regressions with 0.2, 0.4, 0.6, and 0.8 quantiles of mobile check-ins. The solid line connects the parameter estimates of the quantile regressions, with the shaded area being their 95% confidence intervals. We find that the quantile regression parameter estimates increase with quantiles in general. This suggests that the impact of online management responses on check-ins is more pronounced for high quantiles (restaurants with a larger number of check-ins). A plausible explanation is that more customers may visit the Yelp page and observe online management responses for restaurants with a larger number of check-ins, and hence the impact of online management responses is stronger.

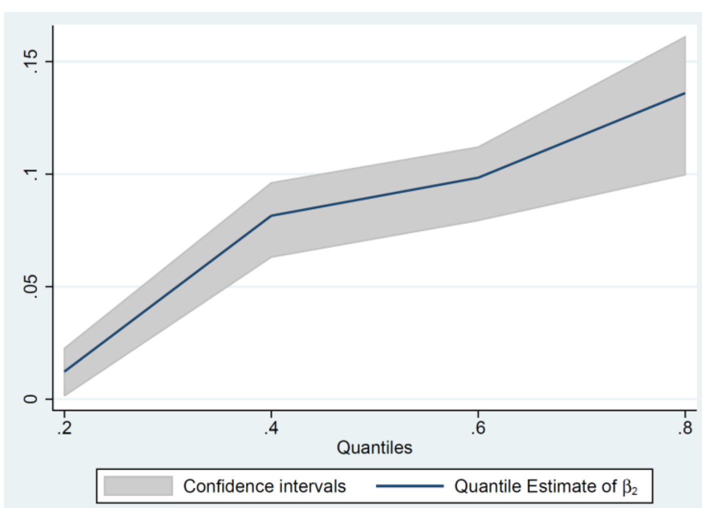

Figure B.1 Quantile Regression Estimates for Regression Model 2

To further investigate the heterogeneity of our results, we split our sample into two subsamples according to the popularity of restaurants (the total number of restaurant reviews): whether or not a restaurant has a top 50-percentile in terms of the total number of restaurant reviews. The estimation results are presented in Table B.1 as follows. We find that the impact of online management responses on check-ins is more pronounced for popular restaurants, which is consistent with our quantile regression results.

Table B.1 Subsample Estimation Results: Different Levels of Restaurant Popularity

	(1)	(2)
VARIABLES	Top 50-Percentile	Bottom 50-Percentile
PostLaunch*OwnerBinaryResp	0.124***	0.0329***
	[3.674]	[2.865]

Robust t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

Online Appendix C: The Moderating Role of the Length of Online Management Responses

In this appendix, we examine the moderating role of the length of online management responses in the following regression equation:

$$\mathbf{log}(checkin_{it}) = c_i + \beta_0 + \beta_1 PostLaunch_t + \beta_2(PostLaunch_t * OwnerBinaryResp_{it}) + \beta_3(PostLaunch_t * OwnerBinaryResp_{it} * LengthResp_{it}) + \beta_4 Controls + \varepsilon_{it}, \quad \text{(C.1)}$$

where the newly constructed variable $LengthResp_{it}$ is the average online management response (text) length for a given month of a restaurant. The regression results are presented in Table C.1. We find that the coefficient on the triple interaction term, β_3, is significantly positive, which implies that a longer online management response has a larger effect on attracting check-ins.

Table C.1 The Impact of Business Owner Responses on Restaurant Performance: The Moderating Role of the Length of Responses

	(1)	(2)
VARIABLES	FE	FE, Robust S.E.
PostLaunch	-0.0109**	-0.0109**
	[-2.144]	[-2.065]
PostLaunch*OwnerBinaryResp	0.0674***	0.0674***
	[4.647]	[3.442]
PostLaunch*OwnerBinaryResp*LengthResp	0.000158***	0.000158***
	[3.215]	[2.836]
Observations	163,757	163,757

t-statistics or robust t-statistics in brackets *** p<0.01, ** p<0.05, * p<0.1

Online Appendix D: The Rounding Thresholds of Displayed Review Ratings

The incentive of writing online management responses may differ when restaurants have different review ratings. In particular, Yelp aggregates all reviews for a given business and displays the average rating prominently. However, when Yelp computes the average rating, they round off to the nearest half star. Two restaurants that have similar average ratings can thus appear to be of very different quality. For example, a restaurant with an average rating of 3.24 displays a 3-star average rating while a restaurant with an average rating of 3.26 displays a 3.5-star average rating (Anderson and Magruder 2012). Therefore, when the review rating of a restaurant is close to the rounding threshold, it has a greater incentive to write responses to consumer comments in order to keep its current star rating (when the review rating is a little bit higher than the threshold) or boost the star rating to the next level (when the review rating is a little bit lower than the threshold).

In the regression equation (D.1), we add a new binary variable indicating whether the average cumulative review rating of a restaurant is close to the rounding thresholds of displayed review rating ($RatingCloseRounding_{it}$). We construct this binary variable carefully in the following steps:

(i) As demonstrated in Anderson and Magruder (2012), the rounding thresholds are 1.25, 1.75, 2.25, 2.75, 3.25, 3.75, 4.25, and 4.75. The variable $CumReviewRating_{it}$ is the average cumulative review rating of a restaurant up until time t. If $|CumReviewRating_{it} - rounding\ threshold| < 0.02$, we say that the average cumulative review rating of the restaurant is close to the rounding thresholds.

(ii) If the average cumulative review rating of the restaurant is close to the rounding thresholds, the binary variable $RatingCloseRounding_{it} = 1$, otherwise, $RatingCloseRounding_{it} = 0$.

$$OwnerResponse_{it} = a_i + \beta_0 + \beta_1 CountRest_i + \beta_2 AveReviewRating_{it} + \beta_3 RatingCloseRounding_{it} + \beta_4 Controls + \varepsilon_{it}. \quad \text{(D.1)}$$

The estimation results are presented in Table D.1. We find that the coefficient of $RatingCloseRounding_{it}$ is significantly positive, which implies that a restaurant is more likely to write responses when the review rating is closer to the rounding threshold. The intuition is that restaurant owners are more likely to worry about their displayed Yelp review ratings when their actual review ratings are closer to the rounding threshold. Therefore, they have a greater incentive to respond to consumer comments.

Table D.1 The Effect of Review Rating Rounding Thresholds on Business Owners' Responses

	(1)	(2)
VARIABLES	Random effects: Total Owner Response Counts by Month as DV, Robust S.E.	Random effects: Binary Owner Response as DV, Robust S.E.
CountRest	0.000147***	3.68e-05***
	[3.243]	[3.824]
AveReviewRating	-0.127***	-0.0548***
	[-13.54]	[-21.74]
Population	-1.64e-07	5.43e-09
	[-0.321]	[0.0315]
MedianAge	-0.00238*	-0.000265
	[-1.886]	[-1.106]
MeanIncome	1.28e-06***	3.18e-07***
	[5.625]	[6.285]
RatingCloseRounding	0.685***	0.0954***
	[2.874]	[3.225]
Category dummies	Yes	Yes
City dummies	Yes	Yes
Monthly dummies	Yes	Yes
Constant	0.584***	0.226***
	[8.142]	[16.84]
Observations	140,206	140,206

t-statistics or robust t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

Online Appendix E: Additional Analyses on Review Rating and Online Management Responses

In this appendix, we conduct additional analyses to confirm that restaurants are more likely to respond to consumer reviews with very low ratings. We estimate the following two regression equations:

$$OwnerResponse_{it} = a_i + \beta_0 + \beta_1 CountRest_i + \beta_2 SumLowRating_{it} + \beta_3 Controls + \varepsilon_{it}, \quad \text{(E.1)}$$

$$OwnerResponse_{it} = a_i + \beta_0 + \beta_1 CountRest_i + \beta_2 FractLowRating_{it} + \beta_3 Controls + \varepsilon_{it}, \quad \text{(E.2)}$$

where $SumLowRating_{it}$ is the number of reviews with low rating (1 and 2 stars) for a given month of a restaurant, and $FractLowRating_{it}$ is the fraction of reviews with low rating (1 and 2 stars) for a given month of a restaurant. The estimation results are presented in Table E.1. We find that the coefficients on $SumLowRating_{it}$ and $FractLowRating_{it}$ are significantly positive, which confirms that restaurants are more likely to respond to consumer reviews with very low ratings.

Table E.1 The Effect of Low Review Rating on Online Management Responses

VARIABLES	(1) FE	(2) FE, Robust S.E.
SumLowRating	0.128***	0.128***
	[3.243]	[3.074]
FractLowRating	4.472***	4.472***
	[4.651]	[4.377]
Observations	140,206	140,206

t-statistics or robust t-statistics in brackets *** p<0.01, ** p<0.05, * p<0.1

Online Appendix F: Additional Analyses on Restaurant Heterogeneity

In this appendix, we look at the heterogeneity across restaurants. According to the prior literature (Auty 1992; Clark and Wood 1998; Astuti and Hanan 2012), older consumers are more loyal to existing restaurants, and younger consumers are more likely to explore new restaurants. In our dataset, we know the age of each restaurant, so we split our sample into two subsamples according to the age of restaurants: whether or not a restaurant has a top 50-percentile in terms of the restaurant age. We expect that old consumers are more likely to visit the top 50-percentile restaurants, and younger consumers are more likely to visit the bottom 50-percentile restaurants.

We re-estimate regression equation (2) using subsamples. The estimation results are presented in Table F.1. We find that the impact of online management responses is greater in the bottom 50-percentile subsample than in the top 50-percentile subsample. Because older consumers might be more likely to stick to existing and more established restaurants (top 50-percentile restaurants), there are two plausible explanations for this finding. The first explanation is that our dependent variable (the number of mobile check-ins) does not account for older consumers' visits (older consumers are less likely to post their mobile check-ins), so the coefficient in the second column (top 50-percentile subsample) is likely to underestimate the true effect. The second explanation is that older consumers are less likely to be active on Yelp and hence are less likely to be affected by online management responses on Yelp. Therefore, the impact of online management responses is weaker in the top 50-percentile subsample.

Table F.1 Subsample Estimation Results: Different Levels of Restaurant Age

	(1)	(2)
VARIABLES	Bottom 50-Percentile	Top 50-Percentile
PostLaunch*OwnerBinaryResp	0.107***	0.0825***
	[3.841]	[3.327]

Robust t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

We also collect the demographic information of each zip-code region in our sample. We focus on the percentage of population whose age is greater than 60 in each zip-code region. We split our sample into two subsamples according to the percentage of population whose age is greater than 60 in zip-code regions: whether a restaurant's zip-code region has a top 50-percentile in terms of the percentage of population whose age is greater than 60. We expect that old consumers are more likely to visit the restaurants in the top 50-percentile subsample, and younger consumers are more likely to visit the restaurants in the bottom 50-percentile subsample. We re-estimate the regression equation (2) using subsamples. The estimation results are presented in Table F.2. Again, we find that the impact of online management responses is greater in the bottom 50-percentile subsample than in the top 50-percentile subsample. Like the case in Table F.1, there are two plausible explanations for our finding in Table F.2. The first explanation is that our dependent variable (the number of mobile check-ins) does not account for older consumers' visits, so the coefficient in the second column (top 50-percentile subsample) is likely to underestimate the true effect. The second explanation is that older consumers are less likely to be active on Yelp and hence are less likely to be affected by online management responses on Yelp. Therefore, the impact of online management responses is weaker in the top 50-percentile subsample.

Table F.2 Subsample Estimation Results: Different Levels of Zip-Code Region Population Age

VARIABLES	(1) Bottom 50-Percentile	(2) Top 50-Percentile
PostLaunch*OwnerBinaryResp	0.116***	0.0782***
	[4.215]	[3.116]

Robust t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

Online Appendix G: Ruling out Pre-Treatment Trends

A potential concern in the DID model is whether there is a heterogeneity in the pre-treatment trends between control and treatment groups (Angrist and Pischke 2008; Greenwood and Wattal 2016). If there is a significant heterogeneity in the pre-treatment trends, it suggests that the pre-treatments may disproportionately affect treated units, as opposed to control units, and the "parallel path" assumption is less likely to be satisfied. In our context, the concern of pre-treatment trends arises because unobserved socio-economic factors in each local region may cause heterogeneity in the pre-treatment trends, and more importantly, the pre-treatment trends could affect restaurants' decisions to respond to consumer comments. For example, a restaurant may monitor its in-store traffic to decide whether to respond to consumer reviews: When the number of mobile check-ins decreases, a restaurant may be more likely to write responses. In this section, we conduct two robustness checks to address this concern and rule out the impact of pre-treatments as an alternative explanation for our results.

In the first check, we follow Angrist and Pischke (2008) to control for time trends in the correlated random trend model:

$$\mathbf{log}(checkin_{it}) = c_i + \beta_0 + g_i t + \beta_1 PostLaunch_t$$
$$+\beta_2(PostLaunch_t * OwnerBinaryResp_{it}) + \beta_3 Controls + \varepsilon_{it}, \qquad \text{(G.1)}$$

where g_i is a restaurant-specific time trend for restaurant i. This specification allows treated and control restaurants to follow different trends in a limited but potentially revealing way. It is worth noting that, in the correlated random trend model, $g_i t$ can be correlated with $OwnerBinaryResp_{it}$ because in the estimation process, $g_i t$ will be cancelled out by first differencing equation (G.1) twice, and our estimation will be unbiased. Therefore, the DID model with correlated random trends is likely to be more robust and convincing (Angrist and Pischke 2008). The estimation results of the correlated random trend model are presented in Table G.1, and we find that the estimated effects of interest are changed little by the inclusion of these trends, which rules out time trends as an alternative explanation for our results.

Table G.1 The Impact of Business Owner Responses on Restaurant Performance: Correlated Random Trend

VARIABLES	(4) Correlated Random Trend
PostLaunch	-0.0118***
	[-2.527]
PostLaunch*OwnerBinaryResp	0.0627***
	[2.628]
AveReviewRating	0.0224***
	[7.325]
SeReviewRating	-0.0412***
	[-12.36]
AveReviewLength	0.000215***
	[11.74]
ReviewCount	0.00252***
	[3.247]
MonthlyAveTemp	0.000208**
	[2.024]
BelowFreezingDay	-0.000213**
	[-2.107]
RainDay	-0.000387***
	[-2.548]
SnowDay	-0.00194**
	[-3.018]
Monthly dummies	Yes

t-statistics or robust t-statistics in brackets *** p<0.01, ** p<0.05, * p<0.1

In the second check, we adopt the following relative time model proposed by Autor (2003):

$$\mathbf{log}(checkin_{it}) = c_i + \beta_0 + \beta_1 PostLaunch_t + \sum_{\tau=0}^{m} \beta_{2,-\tau}\left(PostLaunch_{t-\tau} * OwnerBinaryResp_{i,t-\tau}\right) + \sum_{\tau=1}^{q} \beta_{3,+\tau}\left(PostLaunch_{t+\tau} * OwnerBinaryResp_{i,t+\tau}\right) + \beta_4 Controls + \varepsilon_{it}, \qquad \text{(G.2)}$$

where the sums on the right-hand side allow for m lags (post-treatment effects) and q leads (anticipatory effects). Following Autor (2003), we set m = 4 and q = 2. As argued by Autor (2003) and Angrist and Pischke (2008), the basic idea of the relative time model is in the spirit of Granger causality test (Granger 1969): if the impact of pre-treatment trends is a confounding factor that can affect restaurants' decisions to respond to consumer reviews, we should observe that the past values of $\mathbf{log}(checkin_{it})$ can predict

$PostLaunch_t * OwnerBinaryResp_{it}$ (the in-store traffic affects the decision to respond to consumer comments). On the other hand, if the impact of pre-treatment trends is less of a concern, we should observe that past values of $PostLaunch_t * OwnerBinaryResp_{it}$ can predict $\mathbf{log}(checkin_{it})$ while future values of $PostLaunch_t * OwnerBinaryResp_{it}$ cannot.

The estimation results of the relative time model are presented in Table G.2. We find no significant effects in the two-month period before a restaurant responded to consumer comments, with sharply increasing effects on the number of check-ins in the first few months after responses. These results show that the past values of $PostLaunch_t * OwnerBinaryResp_{it}$ can predict $\mathbf{log}(checkin_{it})$ while future values of $PostLaunch_t * OwnerBinaryResp_{it}$ cannot, which rules out the impact of pre-treatment trends as a confounding factor.

Table G.2 Estimation Results of the Relative Time Model

	(1)	(2)
VARIABLES	FE	FE, Robust S.E.
PostLaunch*OwnerBinaryResp (t + 2)	0.0102	0.0102
	[1.021]	[1.124]
PostLaunch*OwnerBinaryResp (t + 1)	0.0217	0.0217
	[1.254]	[1.207]
PostLaunch*OwnerBinaryResp (t + 0)	0.0536***	0.0536***
	[2.873]	[2.744]
PostLaunch*OwnerBinaryResp (t – 1)	0.0634***	0.0634***
	[3.066]	[2.854]
PostLaunch*OwnerBinaryResp (t – 2)	0.0715***	0.0715***
	[3.267]	[3.102]
PostLaunch*OwnerBinaryResp (t – 3)	0.0513***	0.0513***
	[2.854]	[2.763]
PostLaunch*OwnerBinaryResp (t – 4)	0.0435***	0.0435***
	[2.726]	[2.704]
Monthly dummies	Yes	Yes

t-statistics or robust t-statistics in brackets *** p<0.01, ** p<0.05, * p<0.1

Online Appendix H: Additional Details on Propensity Score Matching

Figure H.1 displays a graphical summary of covariate imbalance showing the standardized percentage bias for each covariate. We can see that the covariate imbalance has been greatly reduced after matching.

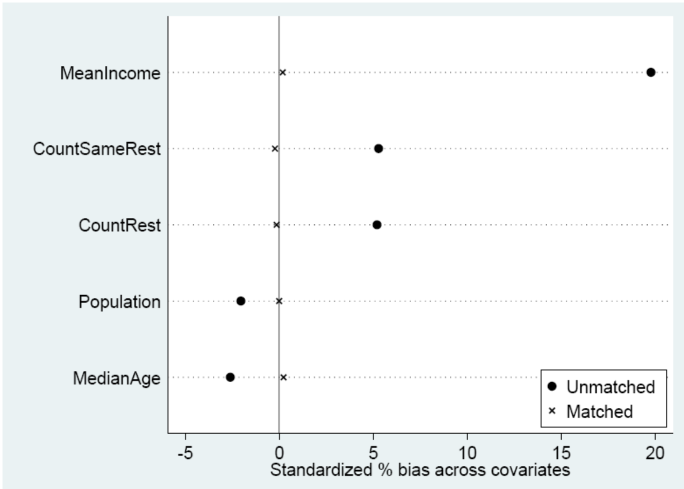

Figure H.1 Standardized Percentage Bias for Each Covariate Before and After Matching