Appendix A Tables

Table A1: Date that intervention and control SMS text messages were sent out over the study period.

For the month of	Date of Reminder SMS	Date of Warning SMS
June 2022	14 July 2022	26 July 2022
July 2022	11 August 2022	31 August 2022
August 2022	9 September 2022	23 September 2022
September 2022	13 October 2022	21 October 2022
October 2022	11 November 2022	22 November 2022
November 2022	14 December 2022	22 December 2022

Note: Tax returns are due on the 15^{th} of the subsequent month. Reminder and warning messages for tax returns of the months June - November 2022 were sent in July - December. Each row reports the date that the reminder and warning SMS text messages were sent out to the study sample for each month.

Table A2: Description of the sample at baseline

Variable	(1) BM+BM	(2) BM+LA	(3) BM+AC	(4) EQ+BM	(5) EQ+LA	(6) EQ+AC	(7) LA+BM	(8) LA+LA	(9) LA+AC	(10) NO SMS	(11) p-value
Dummy: Registered taxpayer	0.976	0.972	0.974	0.976	0.973	0.971	0.976	0.972	0.968	0.972	0.886
is male	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
Dummy: Registered entity	0.214	0.219	0.198	0.216	0.210	0.190	0.214	0.211	0.228	0.198	0.196
is a corporation	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	
Years since enrolled	2.712	2.743	2.713	2.754	2.817	2.757	2.721	2.759	2.651	2.701	0.404
with KPRA	(0.044)	(0.045)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)	(0.043)	
Number of months	3.640	3.621	3.668	3.735	3.610	3.636	3.722	3.610	3.727	3.684	0.791
tax filed Jun-Nov'21	(0.064)	(0.063)	(0.063)	(0.063)	(0.064)	(0.064)	(0.063)	(0.064)	(0.063)	(0.063)	
Avg amount filed	184.431	143.254	46.894	55.366	50.848	67.557	213.068	34.324	136.410	109.443	0.851
Jun-Nov'21 (000's)	(142.598)	(103.616)	(16.323)	(19.277)	(11.209)	(17.682)	(187.655)	(9.786)	(70.527)	(45.314)	
Construction	0.564	0.575	0.567	0.543	0.567	0.572	0.552	0.571	0.559	0.572	0.654
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	
Consultancy Services	0.037	0.027	0.039	0.031	0.038	0.029	0.033	0.025	0.032	0.032	0.309
Hospitality	0.091	0.085	0.091	0.096	0.082	0.069	0.083	0.091	0.085	0.093	0.191
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	
Transportation	0.076	0.086	0.084	0.086	0.087	0.090	0.077	0.077	0.083	0.080	0.831
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	
Agents	0.033	0.030	0.030	0.032	0.031	0.027	0.038	0.029	0.027	0.030	0.803
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	
Withholding agents	0.040	0.048	0.035	0.054	0.035	0.045	0.037	0.048	0.042	0.034	0.027**
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	
Based in provincial capital	0.229	0.203	0.238	0.210	0.226	0.227	0.239	0.228	0.221	0.214	0.171
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	
Observations	1808	1809	1809	1809	1809	1808	1808	1809	1809	1809	

Note: This table provides sample means for variables mentioned in each row by group. For instance, column 1 provides mean for the control ('BM + BM') group. Column 11 provides p-value from a F-test of equality across all control and treatment means. The gender of registered taxpayer is the gender of the authorised representative, which all entities must nominate at the time of registration and, for corporations, may differ from the owner. Gender and years since registration are missing for 47 and 42 participants, respectively. The average amount filed is calculated on a subsample of 13036 participants who filed taxes in the relevant period. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A3 – Description of filers and never-filers at baseline

Variable	(1) Total	(2) Never filers	(3) Filers	(4) *p*-value
Dummy: Registered taxpayer is male	0.973	0.973	0.973	0.998
Dummy: Registered entity is a corporation	0.210	0.257	0.193	0.000***
Years since enrolled with KPRA	2.733	2.733	2.733	0.998
Number of months tax filed June-Nov 2021	3.665	1.064	4.566	0.000***
Average amount filed June-Nov 2021 (000's)	104.122	6.484	113.569	0.264
Service sector: Construction	0.564	0.456	0.602	0.000***
Service sector: Consultancy Services	0.032	0.033	0.032	0.696
Service sector: Hospitality	0.087	0.104	0.081	0.000***
Service sector: Transportation	0.083	0.060	0.091	0.000***
Service sector: Agents	0.031	0.023	0.033	0.000***
Service sector: Withholding agent	0.042	0.161	0.000	0.000***
Dummy: Based in provincial capital (Peshawar)	0.223	0.160	0.245	0.000***
Observations	18,087	4,652	13,435	

Note: This table provides sample means for variables mentioned in each row. Column 1 provides mean values for the total sample, column 2 provides the mean for respondents for respondents who never filed returns within the study period, and column 3 provides the mean for those who filed returns at least once over study period. Column 4 provides p-value from a t-test on differences across means for never and ever filers. Gender and years since registration are missing for 47 and 42 participants, respectively. Average amount filed is calculated on a subsample of 13036 participants who filed taxes in the relevant period. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table A4: Impact on delays (days) past monthly due date

Delays (days) in filing for:	(1) Jun	(2) Jul	(3) Aug	(4) Sep	(5) Oct	(6) Nov
BM+LA	-3.826	-0.821	-1.498	-0.768	-1.121	-2.860
	(2.053)*	(1.992)	(1.869)	(1.719)	(1.562)	(1.372)**
BM+AC	-0.188	1.884	1.508	0.492	1.476	1.779
	(2.168)	(2.047)	(1.945)	(1.721)	(1.625)	(1.504)
EQ+BM	-0.712	-0.767	1.300	1.062	1.354	-0.069
	(2.162)	(1.964)	(1.920)	(1.731)	(1.616)	(1.424)
EQ+LA	0.506	2.359	2.497	2.148	1.914	-0.188
	(2.206)	(2.092)	(1.985)	(1.782)	(1.647)	(1.449)
EQ+AC	-1.417	0.342	0.234	0.126	-1.075	-2.098
	(2.099)	(2.006)	(1.892)	(1.685)	(1.528)	(1.354)
LA+BM	-2.217	-1.790	-1.685	-0.463	-2.073	-1.806
	(2.106)	(1.954)	(1.865)	(1.699)	(1.523)	(1.389)
LA+LA	1.384	2.311	1.320	1.077	0.212	-0.709
	(2.243)	(2.099)	(1.928)	(1.736)	(1.586)	(1.424)
LA+AC	-4.034	-1.920	-2.525	-0.619	-1.635	-1.031
	(1.997)**	(1.883)	(1.776)	(1.644)	(1.483)	(1.357)
NO SMS	-1.384	1.274	1.360	0.585	0.499	-0.340
	(2.165)	(2.092)	(1.971)	(1.735)	(1.601)	(1.417)
Controls	No	No	No	No	No	No
F.E.	No	No	No	No	No	No
Mean(control)	25.958	19.612	17.706	14.045	13.167	11.600
Observations	13092	12589	12501	12344	12116	11949

Note: The table summarizes OLS estimation of equation 1 for the delays past monthly due date in filing for month mentioned in column header: column 1 present results for June, column 2 for July, column 3 for August, column 4 for September, column 5 for October and column 6 for November tax return filed. We have filing data (time and amount) for up to 150 days after the due date for 13,092, 12,589, 12,501, 12,344, 12,116, and 11,949 filers corresponding to the months from June to November, respectively. The pooled sample for these six months includes 18,087 unique filers. Predicted values from the regression are plotted in Figure 3 BM refers to baseline, LA refers to loss aversion message, EQ refers to the equivalent reciprocity message and AC is the active choice message. Reminder messages are labeled first, warning messages type are reported second in the form reminder type + warning type. 'Mean (control)' is the mean outcome value for the control (BM + BM) group. Errors are clustered at the individual level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table A5: Impact on delays (days) past monthly due date, with individual controls and fixed effects

Delays (days) in filing for:	(1) Jun	(2) Jul	(3) Aug	(4) Sep	(5) Oct	(6) Nov
BM+LA	-4.113	-0.765	-1.500	-0.843	-1.362	-3.061
	(2.059)**	(2.000)	(1.874)	(1.722)	(1.563)	(1.376)**
BM+AC	-0.571	1.434	1.402	0.263	1.001	1.544
	(2.158)	(2.036)	(1.947)	(1.717)	(1.613)	(1.497)
EQ+BM	-0.962	-0.906	1.209	0.833	1.057	-0.310
	(2.155)	(1.956)	(1.913)	(1.729)	(1.617)	(1.426)
EQ+LA	0.916	2.865	2.864	2.360	2.226	0.082
	(2.204)	(2.100)	(1.989)	(1.789)	(1.650)	(1.453)
EQ+AC	-1.562	0.351	0.404	0.255	-1.159	-2.101
	(2.095)	(1.999)	(1.888)	(1.677)	(1.522)	(1.353)
LA+BM	-2.047	-1.836	-1.469	-0.444	-2.084	-1.842
	(2.079)	(1.930)	(1.852)	(1.685)	(1.514)	(1.383)
LA+LA	1.519	2.683	1.667	1.396	0.242	-0.718
	(2.244)	(2.092)	(1.925)	(1.731)	(1.580)	(1.420)
LA+AC	-4.294	-2.132	-2.643	-0.814	-1.861	-1.281
	(1.993)**	(1.882)	(1.782)	(1.652)	(1.484)	(1.361)
NO SMS	-1.685	1.219	1.263	0.464	0.338	-0.481
	(2.158)	(2.085)	(1.952)	(1.716)	(1.582)	(1.405)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Mean(control)	25.958	19.612	17.706	14.045	13.167	11.600
Observations	12968	12467	12380	12226	12000	11837

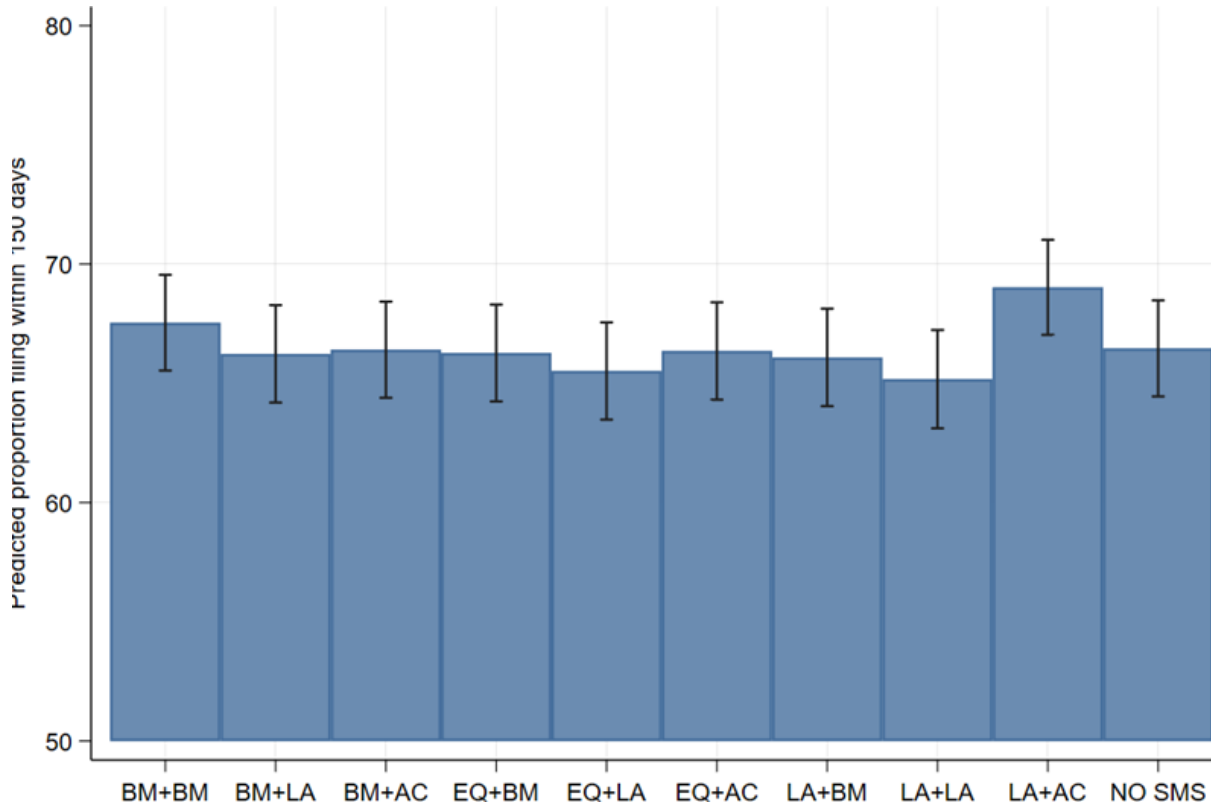
Note: The table summarizes OLS estimation of equation 1 for the delays past monthly due date in filing for month mentioned in column header: column 1 present results for June, column 2 for July, column 3 for August, column 4 for September, column 5 for October and column 6 for November tax return filed. We have filing data (time and amount) for up to 150 days after the due date for 13,092, 12,589, 12,501, 12,344, 12,116, and 11,949 filers corresponding to the months from June to November, respectively. The pooled sample for these six months includes 18,087 unique filers. BM refers to baseline, LA refers to loss aversion message, EQ refers to the equivalent reciprocity message and AC is the active choice message. Reminder messages are labeled first, warning messages type are reported second in the form reminder type + warning type. Controls include indicators for whether the taxpayer represents a corporation and the time since they registered with KPRA. Fixed effects (F.E.) include region and sector fixed effects. 'Mean (control)' is the mean outcome value for the control (BM + BM) group. Errors are clustered at the individual level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Appendix B Figures

Figure B1: Distribution of amount filed for June - November 2021

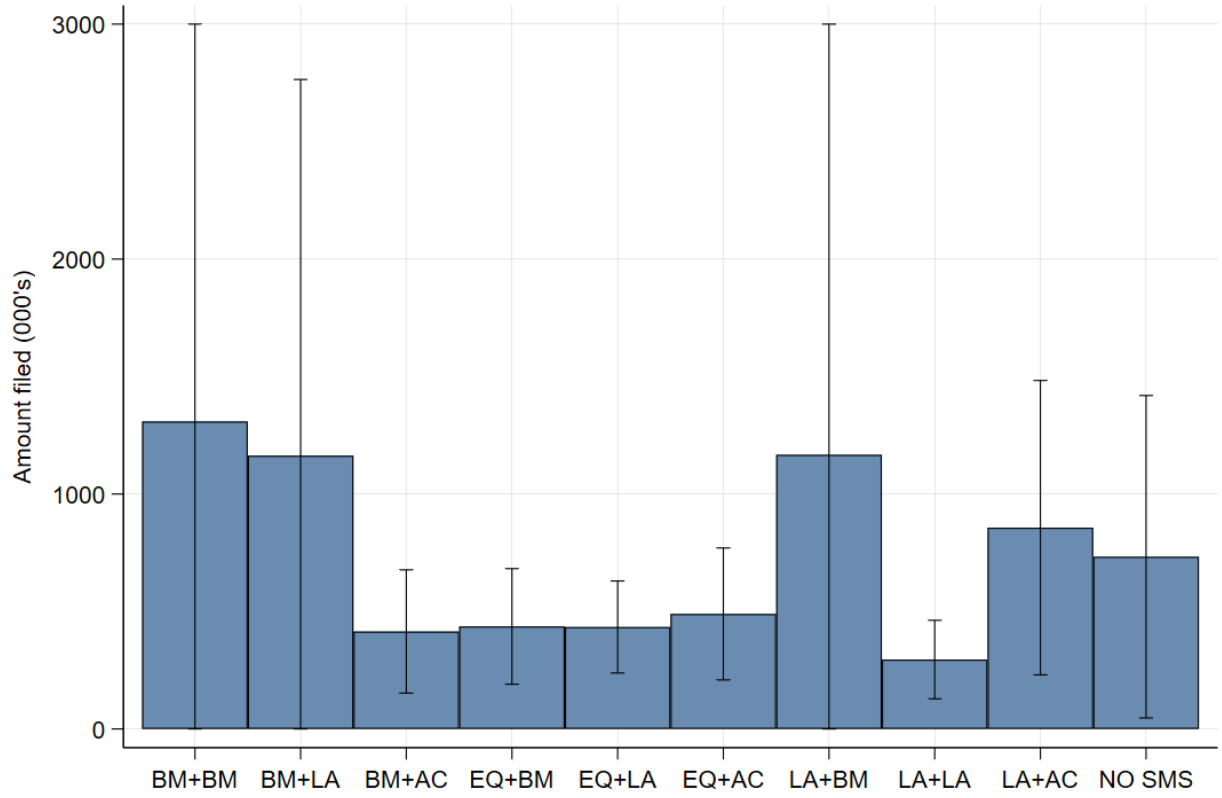
Note: The graph plots the distribution of tax filed for the month mentioned on top of each panel (n = 12,665). The y-axis plots the proportion of the tax-filing sample for each amount specified on the x-axis. 'Null filers' are individuals who have filed tax but have reported zero tax returns.

Figure B2: Average likelihood of filing tax within 5 months of due date for June - November 2022



Note: Bars plot the results of estimating equation 1, obtained from the use of 'margins' command. They represent the predicted probability of filing within 5 months of the due date for each group. Regression results are provided in Table 3, column 1. BM refers to baseline, LA refers to loss aversion message, EQ refers to the equivalent reciprocity message and AC is the active choice message. Reminder messages are labeled first, warning messages type are reported second in the form reminder type + warning type. Vertical lines represent 95% confidence intervals from tests of statistical significance, i.e., difference of each group's predicted values from zero. Stars denote statistical significance of the difference of predicted values of specified intervention group from the predicted values for the control (BM + BM) group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. n = 18087 in both panels.

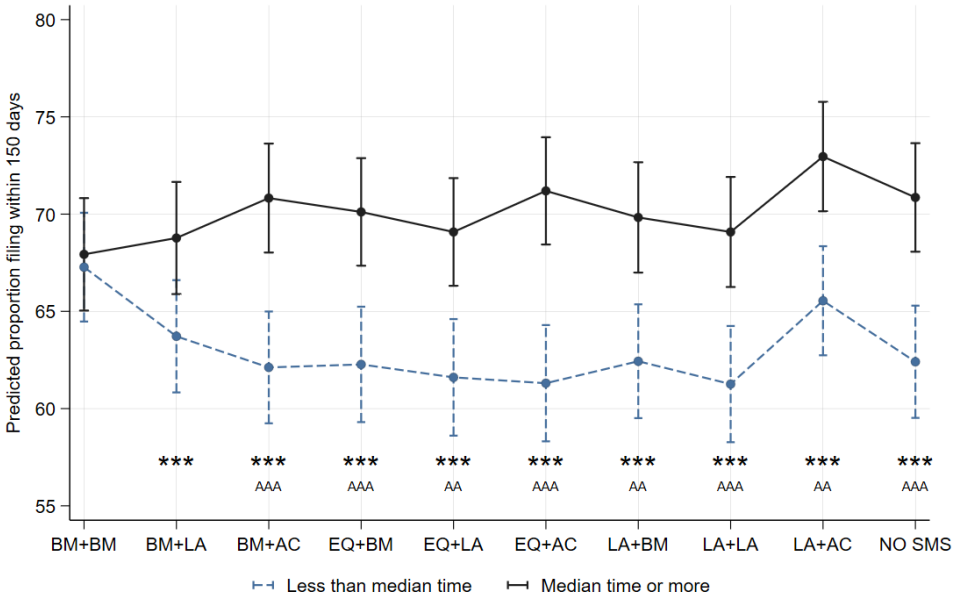
Figure B3: Total tax returns filed June - November 2022



Note: The graph plots predicted amount filed in tax returns, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level. Regression results are provided in appendix Table 3, column 3. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals from tests of statistical significance, i.e., difference of each group's predicted values from zero. Stars denote statistical significance of the difference of predicted values of specified intervention group from the predicted values for the control (BM + BM) group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B4: Heterogeneity in likelihood of filing by early vs recent registration, corporate status and amount filed in the past year

(a) Early vs. Recent registration



(b) Corporations vs. Non-corporate entities

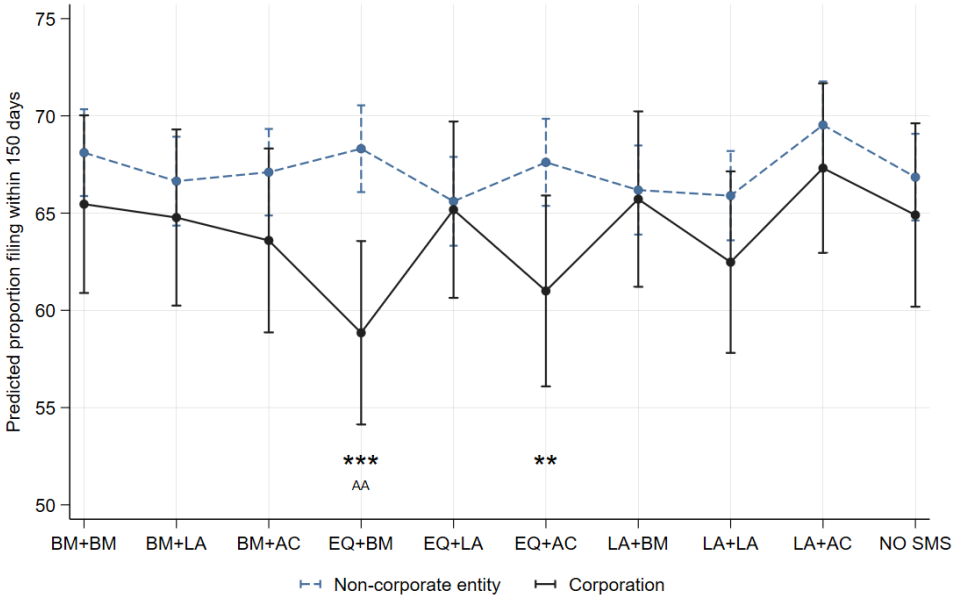
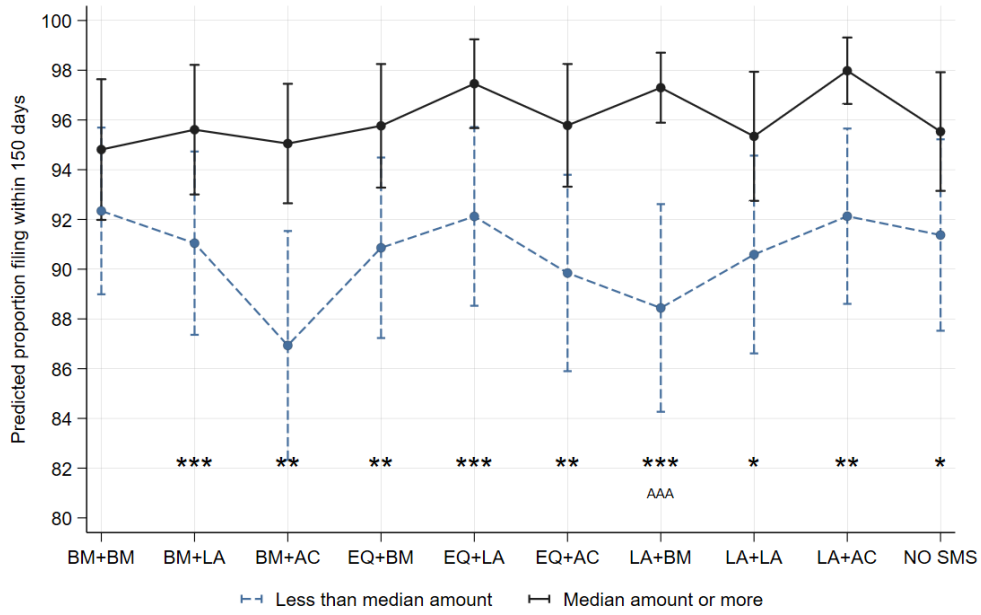


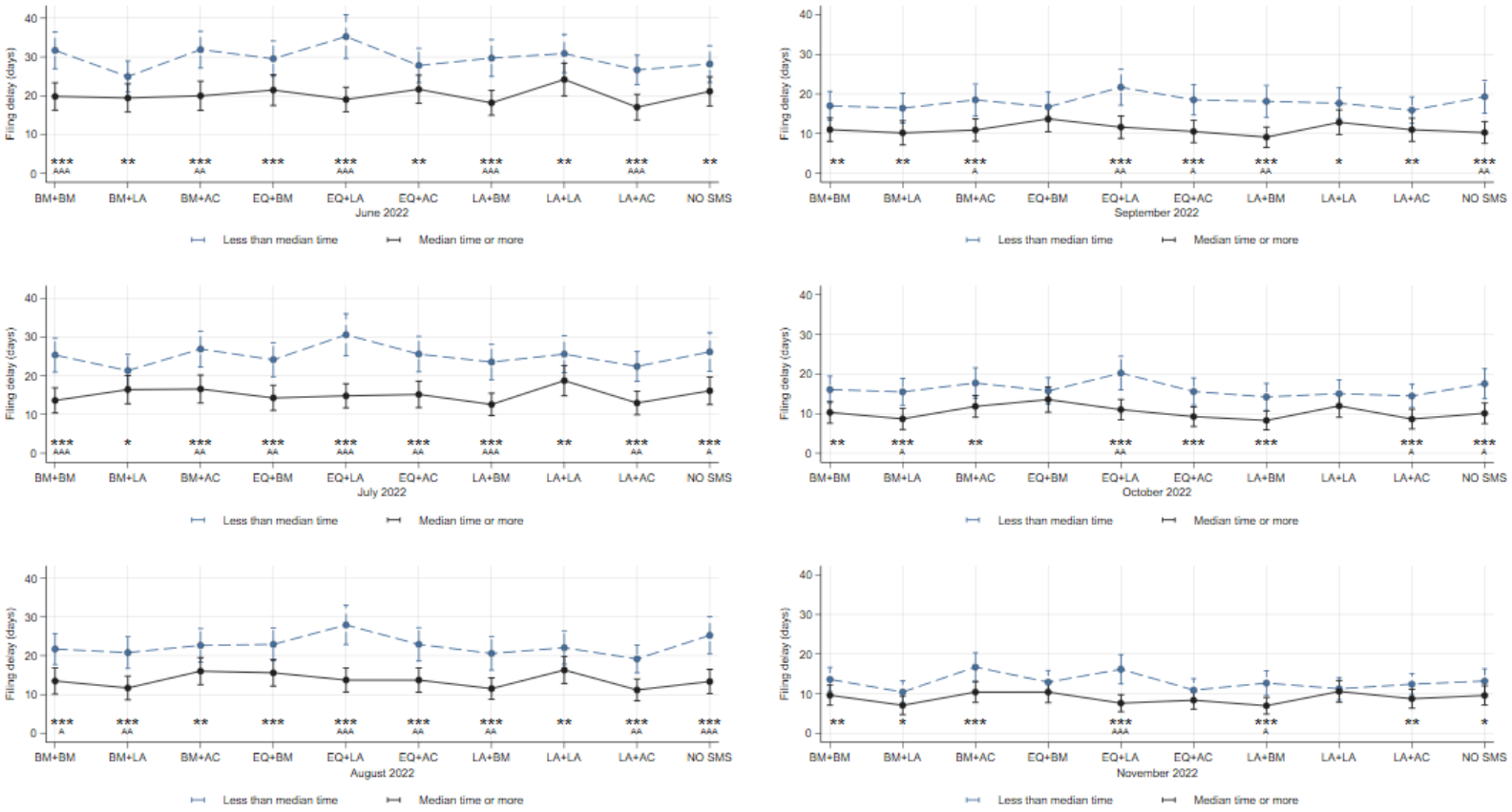
Figure B4 (cont.): Heterogeneity in likelihood of filing by early vs recent registration, corporate status and amount filed in the past year

(c) Large vs. Small filer



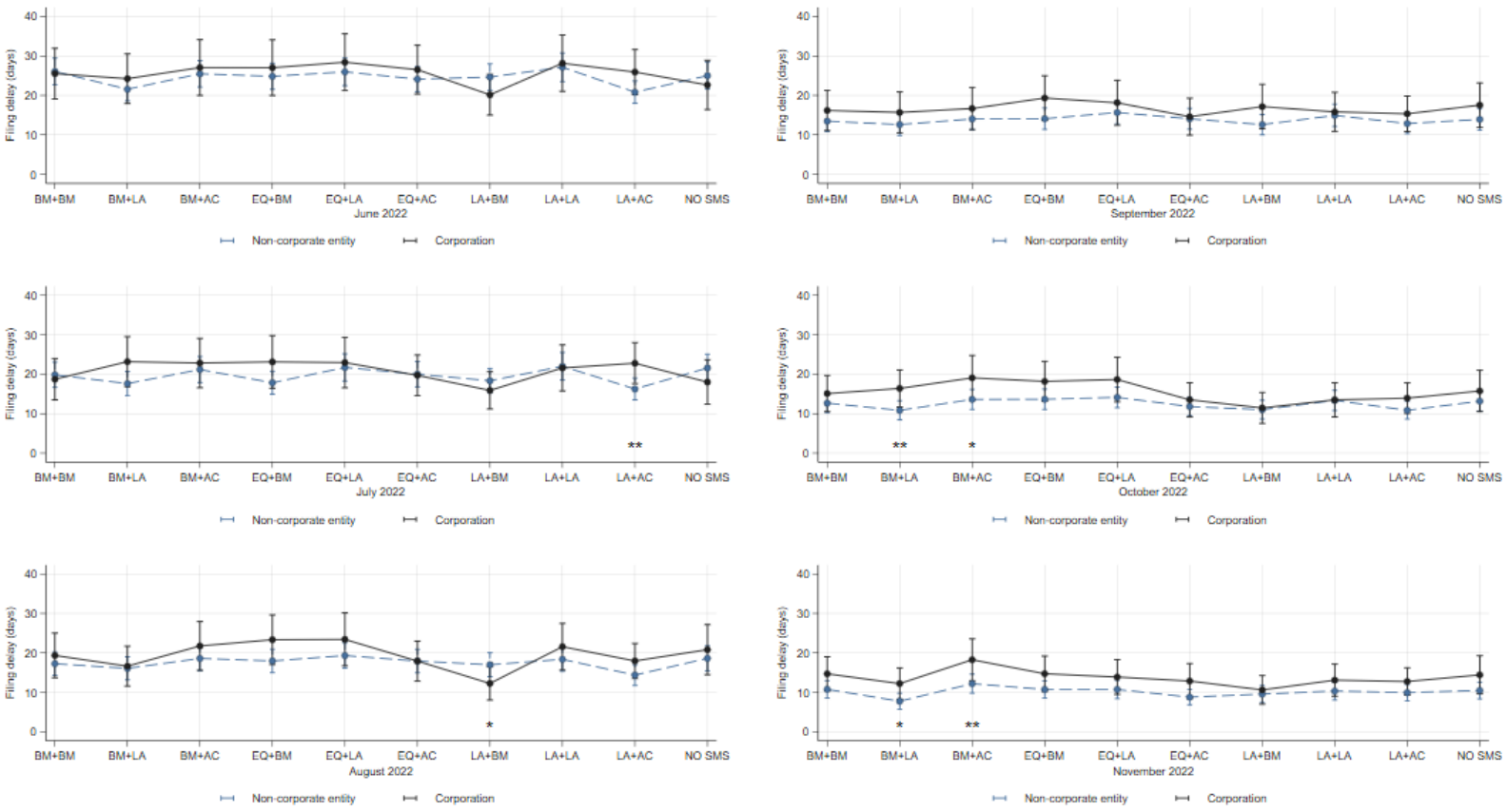
Note: The graph plots the results of estimating equation 1, obtained from the use of 'margins' command. They represent the predicted probability of filing within 5 months of the due date for each group, separately for early and recent registered taxpayers (panel a), corporate and non-corporate entities (panel b), and large and small filer (panel c). n = 18087 in panel (a) and (b) and 3486 in panel (c). BM refers to baseline, LA refers to loss aversion message, EQ refers to the equivalent reciprocity message and AC is the active choice message. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B5: Heterogeneity in days delay in filing (past due date) for each month by early vs recent registration



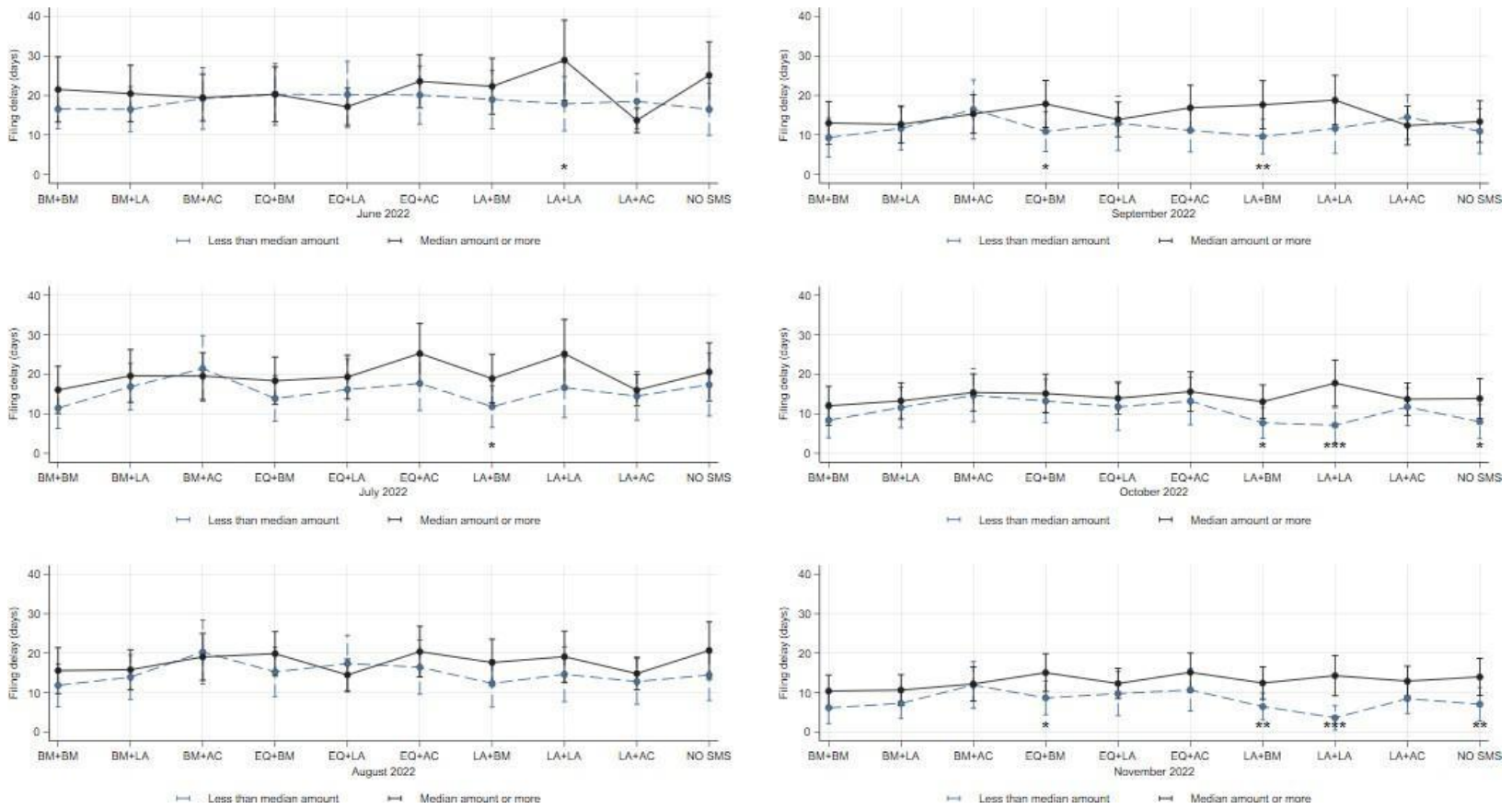
Note: The graph plots predicted (days) delay past the due date in filing taxes, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level, separately for early (black, solid line) and recent (blue, dotted line) registered service providers. Each panel represents results of a regression on delay in filing for the month specified on the bottom. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B6: Heterogeneity in days delay in filing (past due date) for each month by corporations vs non-corporate entities



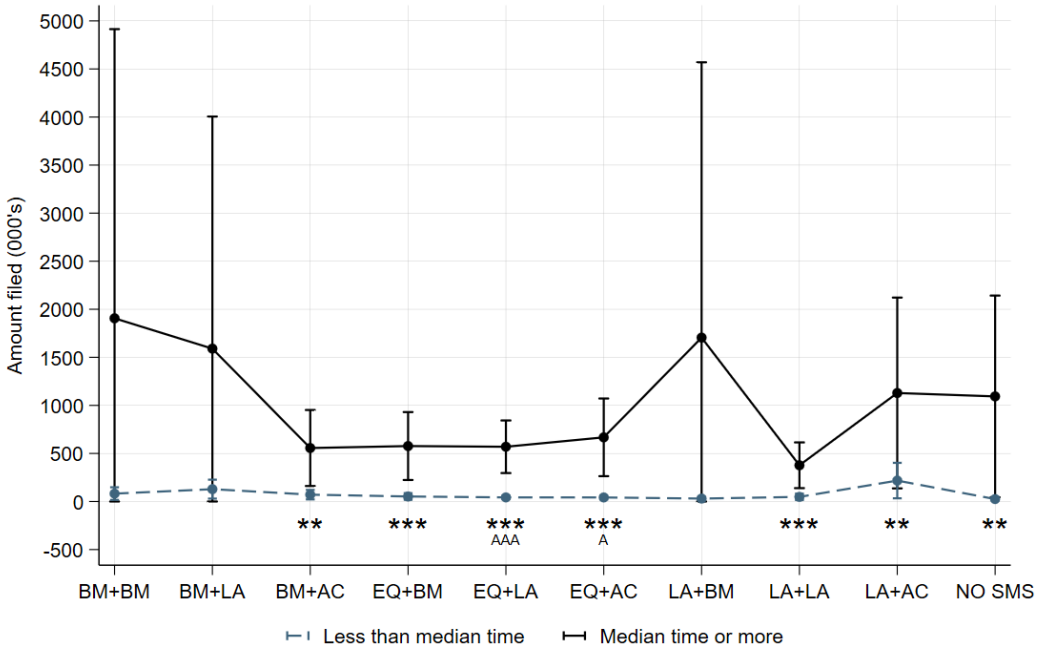
Note: The graph plots predicted (days) delay past the due date in filing taxes, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level, separately for large (black, solid line) and small (blue, dotted line) filers. Each panel represents results of a regression on delay in filing for the month specified on the bottom. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B7: Heterogeneity in days delay in filing (past due date) for each month by large vs small filers



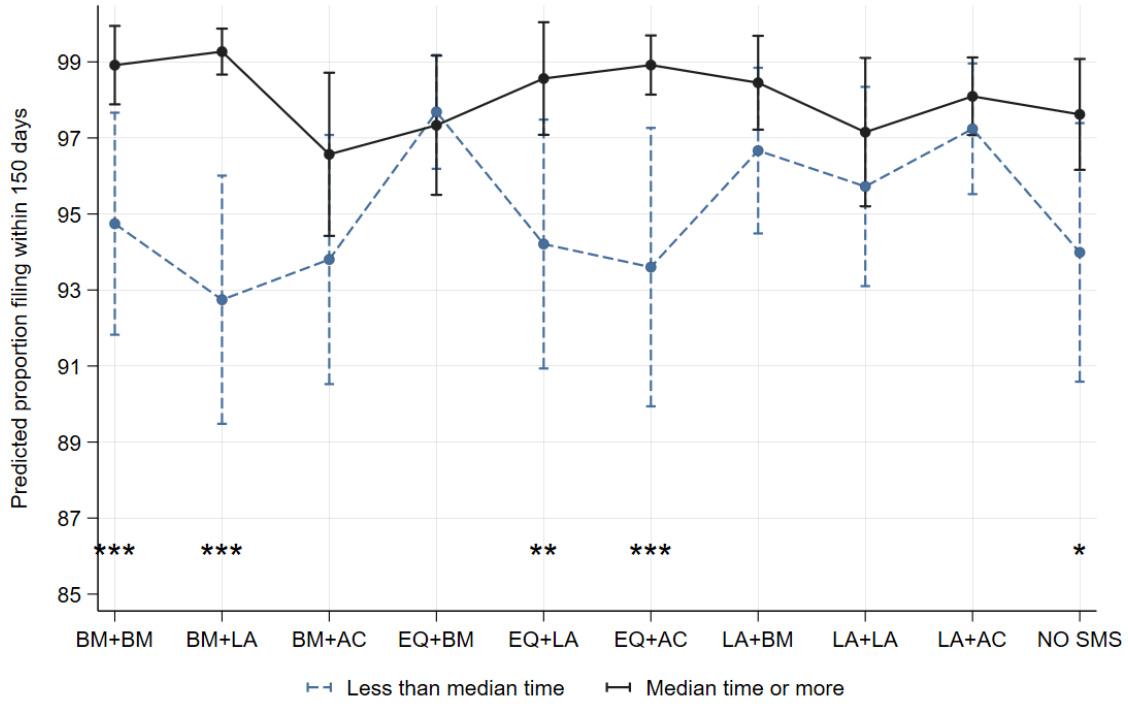
Note: The graph plots predicted (days) delay past the due date in filing taxes, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level, separately for large (black, solid line) and small (blue, dotted line) filers. Each panel represents results of a regression on delay in filing for the month specified on the bottom. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B8: Heterogeneity in total tax returns by early vs recent registration



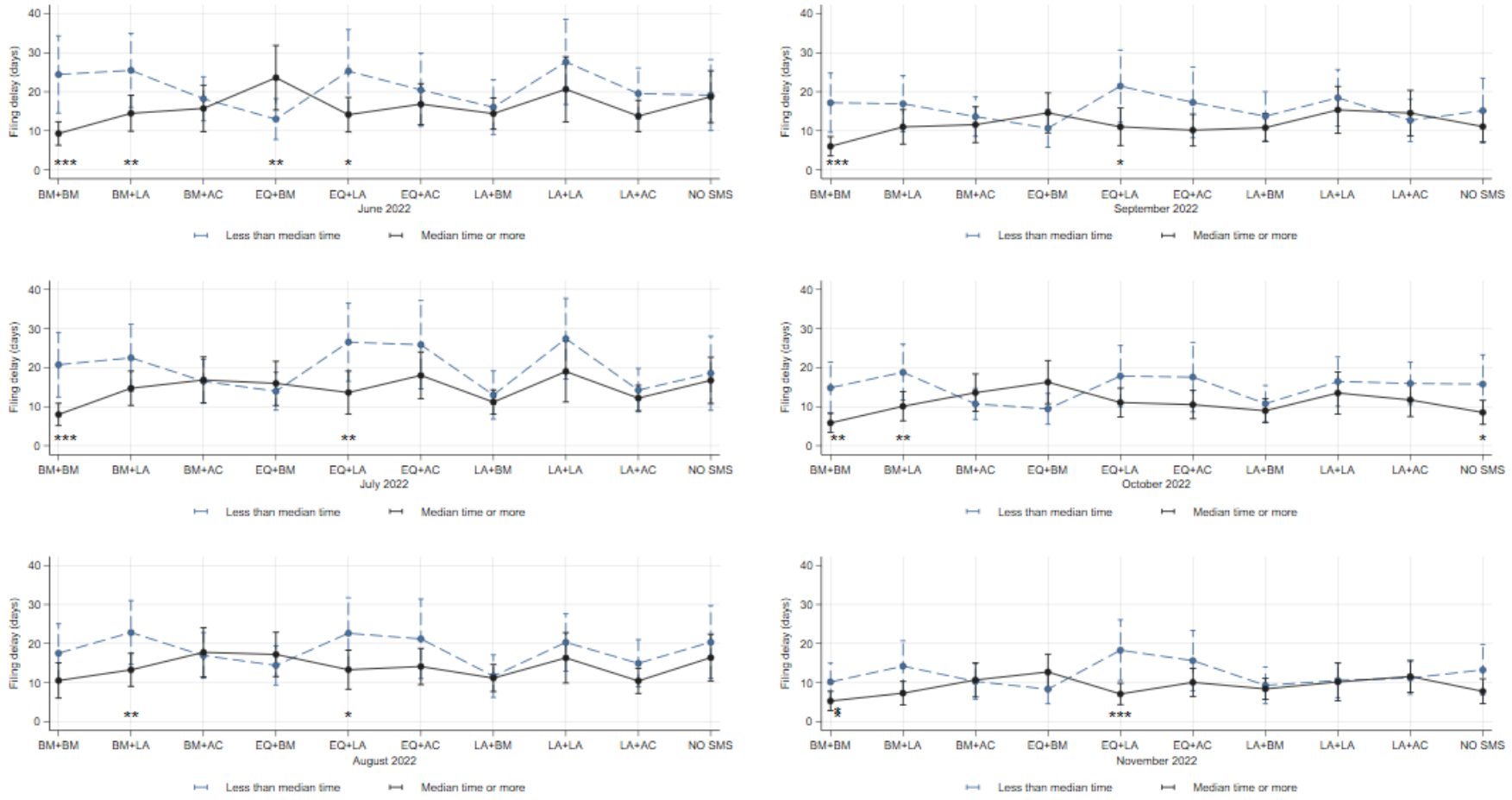
Note: The graph plots predicted amount filed in tax returns, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level, separately for early and recent registered taxpayers. n = 3486. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals for statistical significance. Confidence intervals for the 'Less than median time' sub-sample are also plotted but are very small and close to 0 for most treatment message type. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Bonferroni-Holm adjusted p-values (p_{BH}) are represented with superscript A. ^{AAA} $p_{BH} < 0.01$; ^{AA} $p_{BH} < 0.05$; ^{A} $p_{BH} < 0.1$.

Figure B9: Heterogeneity in likelihood of filing by early vs recent registration (excluding null filers)



Note: The graph plots the results of estimating equation 1, obtained from the use of 'margins' command and replicates the Figure B4 panel (a) excluding respondents who were always null-filers during the 6 months of the study. They represent the predicted probability of filing within 5 months of the due date for each group, separately for early and recent registered taxpayers. n = 2951. BM refers to baseline, LA refers to loss aversion message, EQ refers to the equivalent reciprocity message and AC is the active choice message. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Figure B10: Heterogeneity in days delay in filing (past due date) for each month by early vs recent registration (excluding null filers)



Note: The graph plots predicted (days) delay past the due date in filing taxes, estimated using the 'margins' command and an OLS regression of days delay variable on binary indicators for the treatment message type (including the no SMS group) with errors clustered at the individual level, separately for early (black, solid line) and recent (blue, dotted line) registered service providers. The sample excludes respondents who were always null-filers during the 6 months of the study. n = 2951. Each panel represents results of a regression on delay in filing for the month specified on the bottom. The 'BM + BM' (control) message group is the base group. Vertical lines represent 95% confidence intervals for statistical significance. Stars denote statistical significance of the difference of predicted filing likelihood across taxpayer type for the specified intervention group. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.