ONLINE APPENDIX

Insecurity and Firm Displacement:
Evidence from Afghan Corporate Phone Records

Sylvan Herskowitz
World Bank, Development Impact (DIME)

Joshua E. Blumenstock
University of California, Berkeley

Tarek Ghani
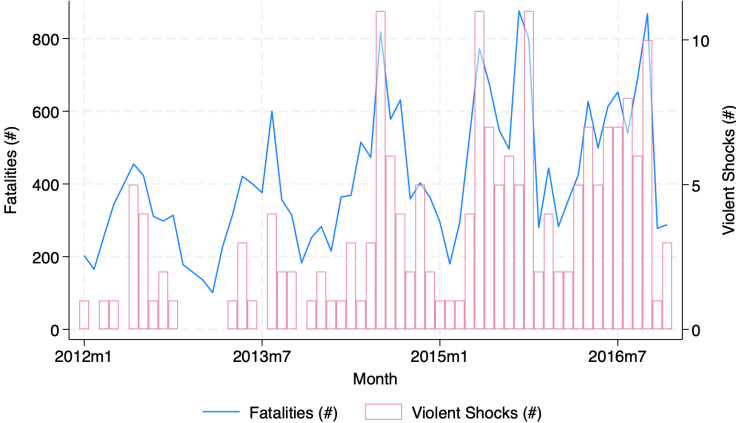Washington University in St. Louis, Olin School of Business

Ethan B. Kapstein
Princeton University and Empirical Studies of Conflict Project

Thomas L. Scherer
University of California, San Diego

Ott Toomet
University of Washington, Information School

A. Online Appendix Figures and Tables

Figure A1: Fatalities and Violent Shocks

Notes: Figure shows confirmed fatalities per month, according to the GTD (solid line, left axis), and the total number of Violent Shocks per month, using the threshold of 23 or more GTD fatalities in a district (bars, right axis).

Figure A2: Spatial Distribution of Violent Shocks
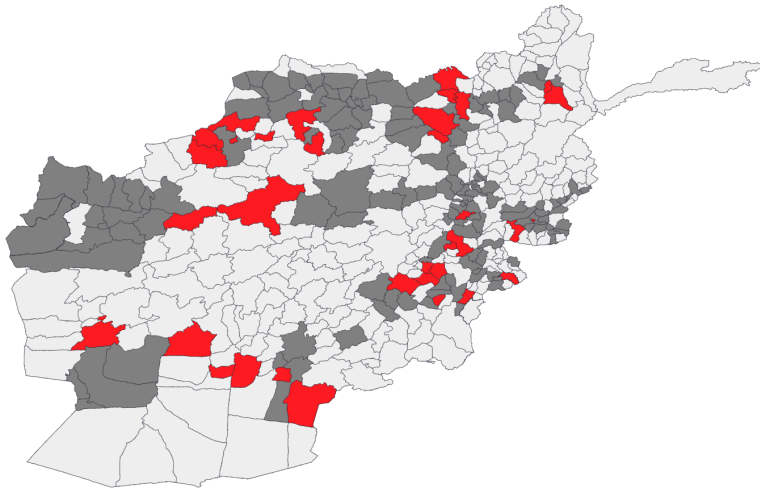
Notes: Figure shows map of Afghanistan. Red-colored districts experienced violent shocks during the study period, while dark grey districts did not. Districts without continuous tower coverage are show in light grey, following the definition in the main estimation sample. (Note that this sample restriction implies more districts are coded light grey here than those without any coverage in the main Figure 1b).

Figure A3: Alternative Violent Shock Thresholds - Any Presence
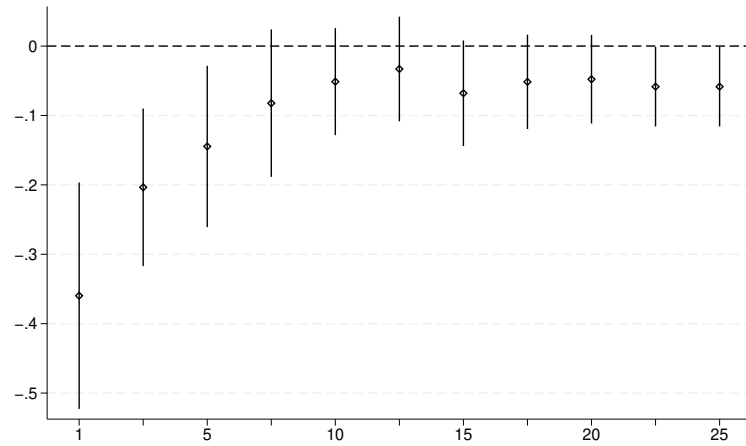
Notes: Figure shows results from a set of estimates varying the threshold of violent shocks used for the lagged measure of the shock. This is defined as being the top x percentile of number of people killed among district-months with positive fatalities. The shock used in the main paper is 6, indicating a district-month in the top 6% of people killed in terrorist attacks which is approximately the top 1% of all observations independent of having any fatalities in terrorist linked attacks in that district-month. The Y-axis shows the estimate of the effect of this shock with the dependent variable of any presence scaled by 100 for readability. 95% confidence intervals are illustrated around the point estimates. The X-axis indicates the percentile threshold of fatalities for a violent shock.

Table A1: Firm Survey Responses

Survey Question	Mean	Std. Dev.	N
Panel A: Business obstacles			
Most important business obstacle - answer includes insecurity (=1)	0.813	0.391	406
Security was an important business obstacle last year (=1)	0.911	0.285	406
Power cuts were an important business obstacle last year (=1)	0.858	0.349	402
Labor problems were an important business obstacle last year (=1)	0.816	0.388	402
Lack of infrastructure an important business obstacle last year (=1)	0.755	0.431	396
Panel B: Concerns about anti-government groups			
Very or extremely affected by insecurity from anti-government groups (=1)	0.784	0.412	403
Very or extremely concerned about land mines and IEDs on roads (=1)	0.851	0.356	403
Very or extremely concerned about attacks with small arms fire (=1)	0.836	0.371	403
Very or extremely concerned about kidnappings (=1)	0.831	0.375	403
Very or extremely concerned about attacks with suicide bombs (=1)	0.93	0.255	402
Panel C: Effects of anti-government groups			
Local employees ever threatened by anti-government groups (=1)	0.226	0.419	403
Local employees ever injured by anti-government groups (=1)	0.082	0.275	402
Local employees ever killed by anti-government groups (=1)	0.052	0.222	405
Firm assets ever threatened or destroyed by anti-government groups (=1)	0.245	0.431	404
Infrastructure ever threatened or destroyed by anti-government groups (=1)	0.58	0.494	402
Panel D: Has your business ever done the following in response to anti-government groups?			
–Spent additional money on private security (=1)	0.457	0.499	403
–Spent money for protection payments (=1)	0.333	0.472	400
–Experienced fall in demand (=1)	0.357	0.48	398
–Delayed an investment in that city or district? (=1)	0.283	0.451	403
–Moved staff away from that city or district (=1)	0.283	0.451	399
–Decreased deliveries to or movement in that city or district? (=1)	0.286	0.453	398
–Changed transportation route? (=1)	0.398	0.49	402
–Changed suppliers to that city or district? (=1)	0.188	0.391	394
–Changed your buyers in that city or district? (=1)	0.151	0.359	397
–Stopped operating in that city or district temporarily? (=1)	0.308	0.462	402
–Stopped operating in that city or district permanently? (=1)	0.075	0.263	402

Notes: Data from original survey of 406 Afghan business owners conducted in 2017, see text for details.

Table A2: Survey Instrument Representativeness Table

	Enterprise Survey (Survey Vars)	CDR Sample (CDR Vars)	CDR Surveyed Sample (CDR Vars)	Survey Sample (Survey Vars)
Num Employees At Present	21.375	52.474	54.788	33.970
Sector Trade (=1)	0.397	0.113	0.103	0.073
Sector Manufacturing (=1)	0.355	0.134	0.379	0.271
Sector Construction (=1)	0.104	0.190	0.185	0.268
Sector Transport (=1)	0.144	0.119	0.106	0.148
Sector Security (=1)	0.000	0.015	0.012	0.010
Sector Finance (=1)	N/A	0.012	0.017	0.033
Sector Information Technology (=1)	N/A	0.005	0.010	N/A
Sector Other (=1)	0.000	0.408	0.187	0.178
HQ in Kabul (=1)	0.404	0.615	0.603	0.700
HQ in Hirat (=1)	0.192	0.167	0.200	0.200
HQ in Balkh (=1)	0.137	0.079	0.103	0.079
HQ in Nangahar (=1)	0.146	0.029	0.025	0.020
HQ in Kandahar (=1)	0.122	0.024	0.012	0.000
HQ in Kunduz (=1)	N/A	0.020	0.012	0.002
N	416	2292	406	406

Notes: Mean values reported for each variable. Enterprise survey means reweighted to reflect nationally representative population. Columns 2 and 3 utilize CDR variables. CDR "Num Employees At Present" calculated based on total MSISDNS for each firm in 2016. CDR sector code was calculated based on a category provided by the phone company, matched to the corresponding two-digit ISIC code (Rev. 4). CDR headquarters are calculated using the firm's first modal district as a proxy. CDR Surveyed refers to the firms in CDR who were surveyed. Columns 1 and 4 utilize survey variables. 'Sectors' and 'Number of Employees at Present' are self-reported, as provided by each survey. World Bank (Enterprise) sector code was calculated based on the four-digit ISIC code (Rev. 3) reported for the primary good or service produced by each firm. Survey headquarters are self-reported, as provided by each survey.

Table A3: Summary Statistics

	Mean	SD	Min	Med	Max
Panel A: Firm Level (N=2,292)					
Total Months Active	32.42	15.05	1.00	41.00	45.00
Total Districts Active	33.75	33.25	1.00	22.00	172.00
Mean Active Districts Per Month	8.53	13.76	0.02	3.60	140.98
Total Employees / Subscribers	33.19	205.84	1.00	4.00	8341.00
Total Calls	94714	813687	1	12279	36102988
Primary Location = Kabul (=1)	0.59	0.49	0.00	1.00	1.00
Primary Location = Provincial Capital (=1)	0.31	0.46	0.00	0.00	1.00
Primary Location = Rural (=1)	0.09	0.28	0.00	0.00	1.00
Active in Primary District (=1)	0.78	0.31	0.02	0.98	1.00
Panel B: District-Month Level (N=7,785)					
Total Firms (Any)	80.90	282.59	0.16	19.15	3934.13
Total Firms (Intense)	36.81	245.10	0.04	3.94	3659.75
Total Employees / Subscribers (Any)	154.81	707.76	0.18	27.22	10351.73
Total Employees / Subscribers (Intense)	82.35	621.23	0.04	5.61	9373.72
Total Calls	11053	84134	3	442	1233656
Total Killed	1.28	5.80	0.00	0.00	244.00
Violent Shock (=1)	0.01	0.09	0.00	0.00	1.00
Insecure District	0.08	0.27	0.00	0.00	1.00
Panel C: Firm-District-Month Level (N=15,809,432)					
Firm Any Activity in District (=1)	0.050	0.22	0.00	0.00	1.00
Firm Any Entry to District (=1)	0.014	0.12	0.00	0.00	1.00
Firm Any Exit from District (=1)	0.015	0.12	0.00	0.00	1.00
Firm Intense Activity In District (=1)	0.017	0.13	0.00	0.00	1.00
Firm Intense Entry to District (=1)	0.002	0.05	0.00	0.00	1.00
Firm Intense Exit from District (=1)	0.003	0.05	0.00	0.00	1.00

Notes: This table shows summary statistics for different levels of aggregation of our dataset of employee mobile phone records. Panel A shows firm-level characteristics, where an 'employee' indicates a mobile subscriber linked to a specific firm's corporate account. Panel B shows district-month level variable,s which include aggregate measures of firm presence and violence (including the two main 'treatment' variables: Violent Shocks and Insecure Province). Panel C provides summary statistics at the level of the firm-district-month, which is the primary unit of observation in our empirical analysis. This is an unbalanced panel of firms (since different firms appear at different points in the panel).

Table A4: Firm District Activity - Alternative Violence Definitions

	(1)	(2)	(3)	(4)	(5)	(6)
	Firm has employee who is active in district (=100)					
Number of Deaths (1 lag)	-0.004***					
	(0.001)					
1-3 Deaths (0-50%)		-0.051***				
		(0.016)				
4-7 Deaths (50-75%)		-0.015				
		(0.028)				
8-22 Deaths (75-95%)		-0.055*				
		(0.031)				
23+ Deaths (>95%)		-0.166***				
		(0.058)				
Kills/100K people			-0.002***			
			(0.001)			
0-3.5 Deaths/100K Pop (0-50%, > 0)				-0.034*		
				(0.018)		
3.5-8.75 Deaths/100K Pop (50-75%, > 0)				-0.057***		
				(0.021)		
8.75-30 Deaths/100K Pop (75-95%, > 0)				-0.058*		
				(0.032)		
>30 Deaths/100K Pop (>95%, > 0)				-0.095		
				(0.061)		
Biggest Event					-0.074	
					(0.048)	
Biggest Two Events						-0.068
						(0.042)
Mean Outcome	4.996	4.996	4.998	4.996	4.996	4.996
Observations	15809432	15809432	15352512	15809432	15809432	15809432
Adjusted R2	0.584	0.584	0.586	0.584	0.584	0.584

Notes: Observation is a firm-district-month. Dependent variable is indicator for any presence, whether any employee linked to the firm made any calls in a given district and month, scaled by 100. All independent variables represent one month lagged measures of violence. Column 1 shows the effects of a continuous number of terrorist attack-linked deaths. Column 2 splits this continuous variable into 4 bins. Column 3 scales the number of people killed by local population size and column 4 divides this per capitized version into bins. Percentage ranges indicate where this level of deaths falls in the distribution of district-months with positive number of deaths. Biggest Event indicates the district-month with the highest number of terrorist-linked deaths for that district whereas Biggest Two Events indicates the two months with the highest number of casualties. Standard errors clustered at district level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A5: Effects of Major Events on Alternative Measures of Firm and Employee Presence

Panel A	(1) Emps-Any	(2) Emps-Intense	(3) Calls	(4) log(Emps-Any)	(5) log(Emps-Intense)	(6) log(Calls)	(7) 2 Calls
Violent Shock (=1)	-0.003	-0.005*	-2.511**	-0.002***	-0.001**	-0.005**	-0.092**
	(0.006)	(0.003)	(1.233)	(0.001)	(0.000)	(0.002)	(0.036)
Mean Outcome	0.250	0.106	13.731	0.058	0.021	0.143	3.256
β / Mean	-0.011	-0.044	-0.183	-0.026	-0.041	-0.038	-0.028
Observations	15809432	15809432	15809432	15809432	15809432	15809432	15809432
Adjusted R2	0.831	0.844	0.821	0.780	0.832	0.778	0.620

Panel B	Emps-Any	Emps-Intense	Calls	log(Emps-Any)	log(Emps-Intense)	log(Calls)	2 Calls
Violent Shock (=1)	0.027	0.033	2.277	0.002	0.003	0.006	0.056
	(0.025)	(0.030)	(2.560)	(0.002)	(0.003)	(0.007)	(0.103)
Violent Shock x Primary	-0.269***	-0.337***	-51.567***	-0.030***	-0.037***	-0.104***	-1.265***
	(0.024)	(0.020)	(3.160)	(0.001)	(0.002)	(0.005)	(0.085)
Mean Non-Primary	0.182	0.049	6.114	0.049	0.014	0.116	2.773
Mean Primary	11.875	9.949	1355.175	1.227	1.137	4.251	74.325
β_1 / Mean Non-Primary	0.150	0.675	0.372	0.036	0.243	0.049	0.020
β_2 / Mean Primary	-0.020	-0.031	-0.036	-0.023	-0.030	-0.023	-0.016
P-value: $\beta_1 + \beta_2 = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	13445387	13445387	13445387	13445387	13445387	13445387	13445387
Adjusted R2	0.927	0.936	0.922	0.809	0.861	0.802	0.643

Notes: This table shows the main effects of violent shocks on firm presence using alternative, intensive measures of firm activity in a given district-month from the CDR. The number of associated employees (emps, subscribers linked to a corporate account) are counted in terms of any or intense presence. Columns 1-3 use the levels of each variable whereas columns 4-6 convert these measures to logs using a transformation of the form Log(1+x) to adjust for skewness of the underlying distribution without dropping zero-valued observations. Emps is a continuous count of the number of unique associated employees who are present in a given district-month for a given firm. The definition of presence is indicated in the column headers following the definitions used throughout the paper. Calls are simply the aggregate number of calls placed from a given district in that month among all affiliated employees for that firm. Column 7 again uses a binary indicator of firm presence but sets the threshold at a minimum of two calls placed from that district by a mobile number linked to that firm in that month. In Panel B, relevant means of the dependent variable are shown for non-primary and primary districts and then scaled effect sizes are shown for shocks in non-primary and primary districts. The reported P-value is a test of shocks in primary districts against zero. All estimates use the full specification from equation 1 while Panel B adds the interaction term of violent shocks and primary location. Standard errors clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Robustness - Work Week, Towers, Provincial Clustering

	(1)	(2)	(3)
Panel A:		Any Presence (=100)	
Violent Shock	-0.109**	-0.116**	-0.143***
	(0.048)	(0.188)	(0.041)
Mean Outcome	4.441	4.499	4.996
β / Mean	-0.025	-0.026	-0.029
Adjusted R2	0.580	0.581	0.584
Panel B:		Intense Presence (=100)	
Violent Shock	-0.063**	-0.072**	-0.087*
	(0.230)	(0.093)	(0.044)
Mean Outcome	1.689	1.532	1.713
β / Mean	-0.038	-0.047	-0.051
Adjusted R2	0.685	0.684	0.687
Clustering	District	District	Province
Panel	Work week	Full	Full
Min Tower Coverage	28 Days	14 Days	28 Days
Observations	15809432	18381651	15809432

Notes: This table shows additional robustness checks to the main specification. Column 1 uses a work week panel where only calls occurring during the work week are used to identify firm presence. Column 2 uses the main paper's coding of firm presence and shows robustness of the main results to a more relaxed constraint of tower day coverage so that districts are only dropped that experience a month with less than 14 days of tower coverage. Column 3 shows robustness to provincial, instead of district, level clustering Each observation is a firm-district-month. All regressions include time fixed effects, district-firm fixed effects, district-season fixed effects, and district linear and quadratic trends. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Heterogeneity: Firm Type

	(1)	(2)	(3)	(4)	(5)
Panel A: Any Presence	Construction	Trade	Manufacturing	Transport	Other
Violent Shock (=1)	-0.364***	0.067	0.041	-0.278*	-0.119**
	(0.134)	(0.121)	(0.145)	(0.163)	(0.054)
Mean Outcome	5.033	4.409	5.172	5.565	4.930
β / Mean	-0.072	0.015	0.008	-0.050	-0.024
Observations	3088396	1874801	2018045	1915456	6912734
Adjusted R2	0.555	0.554	0.590	0.604	0.596
Panel B: Intense Presence	Construction	Trade	Manufacturing	Transport	Other
Violent Shock (=1)	-0.117	0.047	-0.182***	-0.027	-0.100**
	(0.095)	(0.084)	(0.061)	(0.069)	(0.042)
Mean Outcome	1.592	1.290	1.664	1.827	1.864
β / Mean	-0.074	0.037	-0.109	-0.015	-0.054
Observations	3088396	1874801	2018045	1915456	6912734
Adjusted R2	0.667	0.706	0.692	0.683	0.693

Notes: This table shows the effects of violence on firm presence, splitting the sample by different firm industry (type) categories indicated at the top of each column. The estimation uses the specification in equation (1), including time fixed effects, district-firm fixed effects, tower controls, district-season fixed effects, and linear and quadratic trends. Panel A uses an indicator for any firm presence whereas the Panel B uses an indicator for intense firm presence as the outcome.

Table A8: Event Studies

	(1) Any	(2) Intense	(3) Any	(4) Intense
Violent Shock: Lead 6	0.144	-0.043	-0.078	-0.017
	(0.120)	(0.090)	(0.110)	(0.043)
Violent Shock: Lead 5	-0.019	-0.071	-0.077	-0.042
	(0.105)	(0.084)	(0.071)	(0.069)
Violent Shock: Lead 4	0.108	-0.024	0.062	0.006
	(0.105)	(0.061)	(0.075)	(0.047)
Violent Shock: Lead 3	-0.129	-0.204	-0.168*	-0.174
	(0.130)	(0.129)	(0.100)	(0.120)
Violent Shock: Lead 2	0.124*	-0.070	0.084	-0.035
	(0.072)	(0.072)	(0.065)	(0.056)
Violent Shock: Lead 1	-	-	-0.055	-0.031
	-	-	(0.071)	(0.057)
Violent Shock: Current Month	0.040	-0.033	0.011	0.011
	(0.093)	(0.036)	(0.076)	(0.031)
Violent Shock: Lag 1	-0.209*	-0.175**	-0.231**	-0.128**
	(0.114)	(0.084)	(0.094)	(0.064)
Violent Shock: Lag 2	-0.110	-0.146	-0.126	-0.102
	(0.103)	(0.098)	(0.077)	(0.083)
Violent Shock: Lag 3	-0.131	-0.143	-0.128	-0.093
	(0.152)	(0.098)	(0.117)	(0.079)
Violent Shock: Lag 4	-0.060	-0.125**	-0.067	-0.083
	(0.093)	(0.051)	(0.089)	(0.057)
Violent Shock: Lag 5	-0.124	-0.113*	-0.110	-0.059
	(0.104)	(0.061)	(0.081)	(0.073)
Violent Shock: Lag 6	-0.117	-0.120	-0.100	0.025
	(0.162)	(0.073)	(0.114)	(0.087)
Mean Outcome	5.164	1.759	5.164	1.759
F-Test of Leads	0.004	0.461	0.009	0.279
F-Test Sum of Leads	0.629	0.300	0.494	0.388
Observations	13445387	13445387	13445387	13445387
Adjusted-R2	0.590	0.698	0.590	0.698

Notes: Columns 1 and 2 use the event study specification detailed in section 4.3, using the first lead as the reference period. Columns 3 and 4 instead use the time periods before and after the event study window as within firm-district reference periods. "Lead" indicates whether a violent shock happened a given number of months in the future whereas "Lag" indicates violent shocks occurring in the past, with the number indicating the number of months in the future or past. Observation is a firm-district-month. Column headers indicate the measure of firm presence used as the dependent variable. All regressions include time fixed effects, district-firm fixed effects, tower controls, district-season fixed effects, and district linear and quadratic trends. Standard errors clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Entry and Exit - Wider Windows

Dependent Variable:	Entry				Exit			
	Any Presence		Intense Presence		Any Presence		Intense Presence	
Pre Periods	1	3	1	3	1	3	1	3
Post Periods	3	1	3	1	3	1	3	1
Panel A:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Violent Shock (=1)	-0.015	-0.061**	-0.006	-0.028*	0.061	0.068**	0.051	0.030
	(0.020)	(0.029)	(0.014)	(0.015)	(0.037)	(0.032)	(0.034)	(0.019)
Mean Outcome	0.287	0.788	0.086	0.137	0.877	0.310	0.154	0.093
β / Mean	-0.052	-0.077	-0.067	-0.202	0.069	0.219	0.332	0.320
Observations	14620922	14620922	14620922	14620922	14620922	14620922	14620922	14620922
Adjusted R2	0.034	0.033	0.028	0.026	0.034	0.035	0.025	0.028
Panel B:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Violent Shock (=1)	-0.023	-0.083*	-0.003	-0.031	0.048	0.033	0.039	0.018
	(0.024)	(0.043))	(0.011)	(0.023)	(0.038)	(0.024)	(0.027)	(0.018)
Violent Shock x Primary	0.068	0.051	-0.021	0.024	0.238***	0.294***	0.227***	0.175***
	(0.042)	(0.066)	(0.031)	(0.051)	(0.075)	(0.063)	(0.055)	(0.049)
Mean Non-Prim	0.267	0.773	0.074	0.129	0.846	0.295	0.141	0.080
Mean Prim	1.098	0.685	1.319	0.827	1.435	1.833	1.634	2.094
β_1 / Mean Non-Primary	-0.087	-0.108	-0.039	-0.239	0.057	0.111	0.274	0.223
$(\beta_1 + \beta_2)$ / Mean Primary	0.041	-0.047	-0.018	-0.008	0.200	0.178	0.163	0.092
P-value: $\beta_1 + \beta_2 = 0$	0.067	0.374	0.423	0.811	0.000	0.000	0.000	0.000
Observations	12672077	13445387	12672077	13445387	12672077	13445387	12672077	13445387
Adjusted R2	0.034	0.033	0.029	0.026	0.035	0.035	0.027	0.029

Notes: This table generalizes the analysis of firm entry and exit to consider longer periods of time before and after violent shocks. The number of periods considered is indicated, in months, at the top of the table. For example, the outcome measure in column (4) is defined as firm entry into a district (using the intense measure of firm presence) where that firm had NOT been present in any of the 3 previous periods and then WAS present in the current period. By contrast, the outcome in column (7) defines firm exit as occurring when a firm WAS present in the previous period and then WAS NOT present in any of the following three periods. Each observation is a firm-district-month. Binary dependent variables are all scaled by 100 for readability. All regressions include firm-district, time, and district-calendar month fixed effects, as well as district linear and quadratic time trends. Standard errors are clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.

Table A10: Shocks, States, and Primary Locations - Full Interactions

| | *Firm Presence Measure (=100)* | |
| | Any | Intense |
	(1)	(2)
Insecure x Primary	-8.139***	-10.299***
	(0.452)	(0.482)
Insecure x Not Primary	-0.014	-0.044
	(0.078)	(0.036)
Violent Shock x Insecure Dist x Primary	-3.835	-2.155
	(2.670)	(2.979)
Violent Shock x Secure Dist x Primary	-1.116***	-1.630***
	(0.095)	(0.118)
Violent Shock x Insecure Dist x Primary	-0.027	-0.011
	(0.110)	(0.043)
Violent Shock x Secure Dist x Not Primary	-0.026	0.152
	(0.123)	(0.171)
Overall Mean Outcome	4.858	1.684
Mean Insecure Primary	51.507	48.493
Mean Secure Primary	76.446	73.252
Mean Insecure Non-Primary	2.375	0.815
Mean Secure Non-Primary	4.636	1.317
P-value: Primary Equality	0.314	0.861
P-value: Non-Primary Equality	0.993	0.317
Observations	13445387	13445387
Adjusted R2	0.603	0.713

Notes: This table shows the full set of interactions that would build on the results in Table 5. P-values test for equality of coefficients between primary interactions and then non-primary interactions. All regressions include firm-district, time, and district-calendar month fixed effects as well as district linear and quadratic time trends. *** p<0.01, ** p<0.05, * p<0.1.

B. CDR Data Appendix

Our study relies on data from one of Afghanistan's largest private telecommunications operators. This appendix describes the three different sources of information that are used in our empirical analysis, and the main steps in the corresponding data processing. These data do not contain the content of the phone calls and text messages, but only certain metadata about the communication. This includes the parties involved in the communication (anonymized id-s), as well as time and location of the communication. As we treat these data as sensitive and confidential, all personally identifying information was removed prior to our analysis. All research was reviewed and approved by the internal review boards at our respective institutions.

B1. Three Different Data Sources

Call Detail Records The central data source for this analysis is *call detail records* (CDRs). These are datasets, originating from the operator's communication logs, that provide basic information about every call (and text message) in the network. We observe CDRs for 45 months, from April 2013 till December 2016. The most important features in CDRs are date and time, caller's id, receiver's id, and id of the network antenna where the call was initiated (only present for calls). Approximately 250 million calls and a similar number of text messages were conducted in the network each month during the analyzed period. As we do not observe the antenna id for text messages, most of our analysis is based on calls. CDRs are what allows us the deduce the location of every single cellphone over time, given it is used frequently.

Antenna Locations The second and complementary source of information is the spatial location of network antennas. Typically several antennas are attached to a single structure (such as cellphone tower) and we only use the tower location in this study. We have geographic coordinates of 1350 towers, located in 267 districts (out of total 398 districts in Afghanistan). The covered region includes all cities and most of other more densely populated areas (see Figure 1a).

Corporate Subscribers The final related dataset is the list of corporate phones. For each month, the provider lists which phone id's are registered as business phones, and provides basic information about the corresponding businesses. We exclude public and non-profit organizations, such as health, education and media groups, or foreign embassies. We refer to the remaining phones as "corporate subscribers".

As phone numbers occasionally move between different accounts, we disregard numbers that are assigned to multiple businesses, that do not have valid business account id, or that have other irregularities (this amounts to approximately 0.5% of the business phones). Over the observation period, slightly less than 200,000 phones belong to private organizations, out of approximately 10 million distinct phone numbers in data.

This information allows us to distinguish between general call activity and business-related activity. It also permits to assess the size of the firms (in terms of corporate phones), and their geographic and temporal activity patterns. We further categorize the firms into industry-related "segments" based on the operator's internal categorization. The segments are construction, finance, IT and telecommunication, manufacturing and trade, security, transportation, and "other". Note that we cannot use the standard ISIC codes because the operator's internal classification is different.

B2. Constructing Panel Data

Our central empirical approach relies on monthly panel data on firm activity by Afghanistan districts. We count all calls and distinct active subscribers by each firm in each spatio-temporal cell, usually district-month. Based on whether the firm was active in the given cell, we also define it's binary "firm presence" in the cell. We define presence in three different ways:

1. *total activity*, count of all calls and text messages in the relevant district-month.

2. binary *presence* indicator, equal to one if the business had any cellpone activity in the given district-month.

3. *intense presence*, an indicator for district-month where the phone was used most often.

Further, we order the districts according to how many phones have intense presence there. We call the district with the largest presence the "headquarter location". The top 5 districts found in this way show a reasonably good fit with the recorded locations of headquarters and regional offices in other administrative and survey data sources.

Activity distribution shows a prominent right tail while in time, there is no major trends in activity. As expected, Kabul region dominates the the spatial picture but the other major cities are also clearly present. The median value of firm size (phones the firm possesses) is 4, while the mean is 52.26 and the maximum value is 10686.