ONLINE APPENDIX

In this online appendix, we present more historical background and further robustness checks. Table A.1 provides details on the bread enrichments standards proposed by the FDA in 1941. Table A.2 provides more details on the arrival of the boll weevil. In particular, it shows the cumulative percentage of counties in our sample to have been impacted by the boll weevil by year. Most of the invasion occurs between 1917 and 1922 and 22 counties are not treated by the boll weevil.

Table A.3 shows the following. Column (1) uses the log of the pellagra death rate as the dependent variable and uses a different measure of cotton intensity (cotton acres as a percent of all fam acres) than in our baseline regressions. Although the not quite statistically significant (p-value = 0.27), the same sign is obtained as when cotton acres per capita is used. Column (2) uses the log of the pellagra death rate as the dependent variable, but controls the overall death rate (minus pellagra deaths). Our main result remains unchanged. Finally, column (3) uses the ratio of pellagra deaths to overall deaths as the dependent variable and column (4) interpolates pellagra deaths in 1918 and 1919 for NC and uses the log of interpolated pellagra death rate. In columns (3) and (4), the dependent variable decreases after the arrival of the boll weevil and decreases more in high cotton counties.

Table A.4 is requires a more extensive motivation and that is provided in the following section.

Table A.5 uses state level data and presents additional evidence that pellagra declines after a state passes a mandatory fortification law.

IS THERE EVIDENCE OF SELECTIVE MIGRATION?

One key concern with our results centers on migration. While we adopt several remedial strategies in to address concerns about migration and population change, in this section we adopt a series of more direct tests. Our goal here is to explore the possibility that boll-weevil induced out-migration might bias our estimates. Three observations motivate our concerns about migration. First, Lange et al. (2009) show that the boll weevil caused laborers to migrate out of afflicted counties. Second, Ager et al. (2015) show that women in high cotton counties were the most likely to leave the labor force in response to the arrival of the boll weevil. Third, Spark et al. (2015, p.77) show that the most affected groups from pellagra were children age 2-10 and females 22-44. These findings suggest that the two groups most likely to have suffered from pellagra also might have been the most likely to have migrated out of affected counties after the boll weevil arrived. In the face of such selective migration, it is possible that some, or all, of the reductions in pellagra we observe stem not from improved nutrition, but from the elevated rates of outmigration among the groups most likely to have suffered from the disease. Hence, in this section, we run a series of regressions in search of evidence that counties with relatively large numbers of individuals vulnerable to pellagra exhibit elevated rates of outmigration. To measure the extent of such selective out-migration, we again use the data from North Carolina and South Carolina and adopt two separate triple difference strategies similar those employed by Aaronson and Mazmder (2011).

 In our first triple difference strategy, we ask if counties with high initial cotton and high initial pellagra rates exhibit elevated rates of outmigration. More precisely, our regression equation is:

|  |  |  |
| --- | --- | --- |
|  | $$ln[population]\_{ct}=α+γ\*\left[high pellagra\right]\_{c,1915}+β\_{0}\*\left[boll weevil\right]\_{ct}+ γ\*\left[high cotton\right]\_{c,1909}+β\_{1}\*\left[boll weevil\right]\_{ct}×\left[high cotton\right]\_{c}+ β\_{2}\*\left[boll weevil\right]\_{ct}×\left[high pellagra\right]\_{c,1915}+β\_{3}\*\left[boll weevil\right]\_{ct}×\left[high cotton\right]\_{c,1909}×\left[high pellagra\right]\_{c,1915}+θ\_{c}+θ\_{t}+ε\_{ct} $$ | (3) |

All variables are defined as above.

 The main variable of interest for this regression is the triple interaction between boll weevil, high cotton, and high pellagra. If counties with both high initial pellagra and high initial cotton exhibit elevated rates of outmigration, the coefficient on this interaction ($β\_{3})$ would be negative and statistically significant. The coefficients on the boll weevil-high cotton interaction ($β\_{1}$) and the boll weevil-high pellagra interaction ($β\_{2})$ are also of some interest. In particular, the coefficient on the boll weevil-high cotton interaction serves as a sort of replication/verification exercise and asks if the patterns we observe in North Carolina and South Carolina comport with Lange et al. (2009) whose work suggests a negative and significant coefficient (i.e., the boll weevil induced outmigration from high cotton counties). The second interaction on the boll weevil and high pellagra indicators tells us whether places with high pellagra rates (independent of their status in terms of cotton production) are generally observing more outmigration.

 In our second triple difference strategy we focus on women (age 22 to 44) and children (age 2 to 10) and ask if those groups exhibit elevated rates of outmigration in counties with high intensity treatment. Specifically, the regression is as follows:

|  |  |  |
| --- | --- | --- |
|  | $$ln[group\\_pop]\_{gct}=α+\left[group\right]\_{c}+β\_{0}\*\left[boll weevil\right]\_{ct}+ γ\*\left[intensity\right]\_{c}+β\_{1}\*\left[boll weevil\right]\_{ct}×\left[intensity\right]\_{c}+ β\_{2}\*\left[boll weevil\right]\_{ct}×\left[group\right]\_{c}+β\_{3}\*\left[boll weevil\right]\_{ct}×\left[intensity\right]\_{c}×\left[group\right]\_{c}+θ\_{c}+θ\_{t}+ε\_{ct} $$ | (4) |

In equation (4), $ln[group\\_pop]\_{cgt}$ is the population for group *g* in county *c* in census year *t*.

$\left[boll weevil\right]\_{ct}$ and $\left[intensity\right]\_{c}$ have the same definitions as in equation (1), and as above we include county and year fixed effects. Groups are defined by their vulnerability. In one set of regressions, the vulnerable group will be children, age 2 to 10, and in another set of regressions the vulnerable group will be females, age 22 to 44. The non-vulnerable group consists of everyone not contained in the vulnerable group. Hence, the variable $\left[group\right]\_{c}$ is an indicator that equals 1 if the population group *g* is vulnerable, and 0 otherwise.

 The main variable of interest here is the triple interaction between boll weevil, intensity, and group. If women and children (the groups most vulnerable to pellagra) exhibit elevated rates of outmigration the coefficient on this interaction ($β\_{3})$ would be negative and statistically significant.

 The results are reported in Table A.4. The results from column (1) are consistent with Lange et al (2009) (high cotton counties exhibit higher rates of outmigration post boll weevil), and show that high pellagra is not associated with elevated rates of migration in either the double or triple difference. The results from columns (2) and (3) indicate that females (age 22 to 44) do not exhibit significantly higher rates of outmigration post-boll weevil than other (less vulnerable) groups. Similarly, the results from columns (4) and (5) indicated that children (age 2 to 10) do not exhibit significantly higher rates of outmigration post-boll weevil than other (less vulnerable) groups. Overall the patterns observed here provide little support for the hypothesis that the groups most vulnerable to pellagra are migrating out of high cotton counties in response to the boll weevil at higher rates than other groups.

|  |
| --- |
| Table A1: Bread enrichment standards proposed by FDA (1941) |
|  |  |  |  |
|  |  | Minimum (mg) | Maximum (mg) |
|  |  |  |  |
| Thiamine |  | 1.1 | 1.8 |
| Riboflavin |  | 0.7 | 1.6 |
| Niacin |  | 10 | 15 |
| Iron |  | 8 | 12.5 |
| Optional Ingredients: |
| Vitamin D |  | 150 | 750 |
| Calcium |  | 300 | 800 |
|   |   |   |   |
| **Source:** Food and Bread Enrichment 1949 - 1950; National Research Council Committee on Cereals. |

|  |
| --- |
| Table A2: Arrival of the boll weevil |
|  |  |  |
|  | Number of counties | Cumulative percent of sample |
| Year of boll weevil arrival: |  |  |
| 1917 | 1 | 0.71 |
|  |  |  |
| 1918 | 7 | 5.67 |
|  |  |  |
| 1919 | 36 | 31.21 |
|  |  |  |
| 1920 | 13 | 40.43 |
|  |  |  |
| 1921 | 29 | 60.99 |
|  |  |  |
| 1922 | 33 | 84.40 |
|  |  |  |
| Does not arrive | 22 | 100 |
| *Source:* The year that the boll weevil first arrived in a county comes from Lange et al. (2009). |

|  |
| --- |
| Table A3: Additional robustness checks |
|  |  |  |  |  |
|  | Different measure of cotton intensity | Control for overall death rate (minus pellagra deaths) | Dependent variable: Pellagra deaths to overall deaths ratio | Dependent variable: Interpolated log pellagra death rate |
|  | (1) | (2) | (3) | (4) |
|  |  |  |  |  |
| Post boll weevil | -0.202\*\*\* | -0.184\*\*\* | -0.00414\*\*\* | -0.179\*\*\* |
|  | (0.0501) | (0.0505) | (0.00119) | (0.0495) |
|  |  |  |  |  |
| Post boll weevil \* high county pre-boll weevil cotton acres as percent of farm acres (1909) | -0.0777 |  |  |  |
| (0.0702) |  |  |  |
|  |  |  |  |  |
| Post boll weevil \* high county pre-boll weevil cotton acres per capita (1909) |  | -0.161\*\* | -0.00600\*\*\* | -0.151\*\* |
|  | (0.0653) | (0.00187) | (0.0653) |
|  |  |  |  |  |
| Log(overall death rate minus pellagra deaths) |  | 0.111 |  |  |
|  | (0.110) |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| County FE | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes |
| Malaria and urbanization controls | Yes | Yes | Yes | Yes |
|  |  |  |  |  |
| Observations | 1312 | 1312 | 1312 | 1508 |
| Counties | 141 | 141 | 141 | 141 |
| *Notes:* This table reports OLS estimates from equation (3) in the text. The unit of observation is a county-year cell. Standard errors, reported in parentheses, are clustered at the county-level. The variable “high county pre-boll weevil cotton acres as percent of farm acres (1909)” is an indicator if a county was in the top 25 percent of the distribution of cotton acres as a percent of total farm acres 1909. The variable “high county pre-boll weevil cotton acres per capita” is an indicator if a county was in the top 25 percent of the distribution of cotton acres per capita in 1909. Malaria and urbanization controls include the malaria death rate in 1915 and the percentage of the county population that lived in an urban designated area in 1910 both interacted with a full set of year dummies.*Sources:* The dependent variables come from the state health reports of North Carolina and South Carolina. The year that the boll weevil first arrived in a county comes from Lange et al. (2009). The high cotton acres indicator comes from Haines et al. (2015) *United States Agriculture Data, 1840-2010*.\* p<0.1, \*\* p<0.05, \*\*\* p<0.01 |

|  |
| --- |
| Table A4: The boll weevil and migration |
|  |  |  |  |  |  |
|  | (1) | (2) | (3) | (4) | (5) |
| Census years: | 1910, 1920, 1930 |
| Dependent variable: | log(population) | log(female age 22-44 population) or log(population – female age 22-44 population) | log(children age 2-10 population) or log(population – children age 22-44 population) |
| Population: | Entire | Female age 22-44 | Female age 22-44 | Children age 2-10 | Children age 2-10 |
| Intensity1: | High cotton | High cotton | High pellagra | High cotton | High pellagra |
| Intensity2: | High pellagra | None | None | None | None |
|  |  |  |  |  |  |
|  | Coefficients |
|  |  |  |  |  |  |
| β0 = Post BW | 0.0616\*\* | 0.0541\*\* | 0.0130 | 0.0988\*\*\* | 0.0364 |
|  | (0.0268) | (0.0217) | (0.0204) | (0.0249) | (0.0223) |
|  |  |  |  |  |  |
| β1 = Post BW\*Intensity1 | -0.194\*\*\* | -0.165\*\*\* | 0.0128 | -0.187\*\*\* | -0.00646 |
| (0.0616) | (0.0428) | (0.0477) | (0.0425) | (0.0446) |
|  |  |  |  |  |  |
| β2 = Post BW\*Population/Intensity2 | 0.0429 | 0.0393\*\*\* | 0.0203\*\* | -0.0577\*\*\* | -0.0288\*\*\* |
| (0.0665) | (0.0119) | (0.00799) | (0.0122) | (0.00938) |
|  |  |  |  |  |  |
| β3 = Post BW \* Intensity1\*Population/Intensity 2 | 0.0195 | -0.0190 | -0.00879 | 0.0183 | 0.00516 |
| (0.0997) | (0.0133) | (0.0138) | (0.0170) | (0.0184) |
|  |  |  |  |  |  |
|  | Differences (Pre BW - post BW) |
|  |  |  |  |  |  |
| β0 | 0.0616\*\* | 0.0541\*\* | 0.0130 | 0.0988\*\*\* | 0.0364 |
|  | (0.0268) | (0.0217) | (0.0204) | (0.0249) | (0.0223) |
|  |  |  |  |  |  |
| β0 + β1 | -0.133\*\* | -0.111\*\*\* | 0.0259 | -0.0884\*\* | 0.0299 |
|  | (0.0551) | (0.0360) | (0.0412) | (0.0345) | (0.0378) |
|  |  |  |  |  |  |
| β0 + β2 | 0.105\* | 0.0934\*\*\* | 0.0333 | 0.0411\* | 0.00760 |
|  | (0.0622) | (0.0268) | (0.0238) | (0.0228) | (0.0209) |
|  |  |  |  |  |  |
| β0 + β1 + β2 + β3 | -0.0702 | -0.0911\*\*\* | 0.0373 | -0.128\*\*\* | 0.00630 |
|  | (0.0539) | (0.0353) | (0.0403) | (0.0394) | (0.0448) |
|  |  |  |  |  |  |
|  | Difference in Difference |
|  |  |  |  |  |  |
| β1 | -0.194\*\*\* | -0.165\*\*\* | 0.0128 | -0.187\*\*\* | -0.00646 |
|  | (0.0616) | (0.0428) | (0.0477) | (0.0425) | (0.0446) |
|  |  |  |  |  |  |
| β1 + β3 | -0.175\*\* | -0.184\*\*\* | 0.00404 | -0.169\*\*\* | -0.00130 |
|  | (0.0813) | (0.0445) | (0.0477) | (0.0475) | (0.0515) |
|  |  |  |  |  |  |
| β2 | 0.0429 | 0.0393\*\*\* | 0.0203\*\* | -0.0577\*\*\* | -0.0288\*\*\* |
|  | (0.0665) | (0.0119) | (0.00799) | (0.0122) | (0.00938) |
|  |  |  |  |  |  |
| β2 + β3 | 0.0624 | 0.0203\*\*\* | 0.0115 | -0.0394\*\*\* | -0.0236 |
|  | (0.0744) | (0.00598) | (0.0112) | (0.0119) | (0.0159) |
|  |  |  |  |  |  |
|  | Triple difference |
|  |  |  |  |  |  |
| β3 | 0.0195 | -0.0190 | -0.00879 | 0.0183 | 0.00516 |
|  | (0.0997) | (0.0133) | (0.0138) | (0.0170) | (0.0184) |
|  |  |  |  |  |  |
| County FE | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes |
|  |  |  |  |  |  |
| Observations | 423 | 846 | 846 | 846 | 846 |
| Counties | 141 | 141 | 141 | 141 | 141 |
| *Notes:* This table reports OLS estimates. The unit of observation in column (1) is a county-year cell. The unit of observation in columns (2)-(5) is a county-year-population group cell. Standard errors, reported in parentheses, are clustered at the county-level. Column (1) reports the coefficients from a triple-difference estimation where an indicator for Post BW is interacted with a high pellagra and high cotton indicator. A county is considered a high pellagra county if it was in the top 25 percent of the distribution of pellagra death rates in North Carolina and South Carolina in 1915 and 1916. A county is considered a high pellagra cotton if it was in the top 25 percent of the distribution of cotton acres per capita in 1909. In columns (2)-(5) each county has two observations per year. The first observation is the population of a group that was especially susceptible to pellagra; women 22-44 years of age or children 2-10 years of age. The second observation is the population of the rest of the county that is not a part of the susceptible group. The coefficients in columns (2)-(5) are from a triple-difference estimation where an indicator for Post BW is interacted with a high pellagra or high cotton indicator and an indicator if the observation is for the susceptible population.\* p<0.1, \*\* p<0.05, \*\*\* p<0.01 |

|  |
| --- |
| Table A5: Discovery of niacin and fortification laws - state |
|  |  |  |  |  |  |  |  |
|  |  | log pellagra death rate |
|  | (1934-1941) | (1938-1949) | (1934-1949) |
|   | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|  |  |  |  |  |  |  |  |
| Post niacin discovery \* high pre-niacin pellagra death rate (1928) | -0.00269 |  |  |  |  | -0.00900 |  |
| (0.0149) |  |  |  |  | (0.0148) |  |
|  |  |  |  |  |  |  |  |
| Post niacin disocvery \* high pre-niacin cotton acres per capita (1909) |  | 0.0180 |  |  |  |  | 0.0168 |
|  | (0.0121) |  |  |  |  | (0.0128) |
|  |  |  |  |  |  |  |  |
| Post fortification law |  |  | -0.00868\*\* | -0.000822 | -0.00295 | 0.000899 | -0.00477 |
|  |  |  | (0.00343) | (0.00549) | (0.00497) | (0.00659) | (0.00469) |
|  |  |  |  |  |  |  |  |
| Post fortification law \* high pre-niacin pellagra death rate (1928) |  |  |  | -0.0126 |  | -0.0169\* |  |
|  |  |  | (0.00797) |  | (0.00858) |  |
|  |  |  |  |  |  |  |  |
| Post fortification law \* high pre-niacin cotton acres per capita (1909) |  |  |  |  | -0.00892 |  | -0.0126\* |
|  |  |  |  | (0.00793) |  | (0.00692) |
|  |  |  |  |  |  |  |  |
| County FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Other controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
|  |  |  |  |  |  |  |  |
| Observations | 96 | 96 | 144 | 144 | 144 | 192 | 192 |
| States | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| *Notes:* This table reports OLS estimates from equation (3) in the text. The unit of observation is a state-year cell. Standard errors, reported in parentheses, are clustered at the state-level. States includes in the analysis are: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, and Virginia. “Post niacin discovery” is an indicator variable that takes a value of one beginning in 1937 with the discovery of niacin. “Post fortification law” is an indicator variable that takes a value of one beginning in the year that a state past a mandatory fortification law. Alabama passed a law in 1943, Arkansas in 1945, Georgia in 1945, Louisiana in 1942, Mississippi in 1945, North Carolina in 1945, Oklahoma in 1947, and South Carolina in 1942. The variable “high pre-niacin pellagra death rate (1928)” is an indicator if a state was in the top 50 percent of the distribution of pellagra death rates in 1928. The variable “high pre-niacin cotton acres per capita (1909)” is an indicator if a state was in the top 50 percent of the distribution of cotton acres per capita in 1909. Malaria and urbanization controls include the state malaria death rate in 1928 and the percentage of the state’s population that lived in an urban designated area in 1910 both interacted with a full set of year dummies.*Sources:* The dependent variable comes from the *Mortality Statistics of the United States, 1915-1936* and the *Vital Statistics of the United States, 1937-1949*. Cotton acres harvested were taken from the United States Department of Agriculture’s National Agricultural Statistics Service Database (Quick Stats 2.0). \* p<0.1, \*\* p<0.05, \*\*\* p<0.01 |