Estimating the prevalence of the COVID-19 infection, with an application to Italy

Franco Peracchi
Georgetown U., EIEF & U. Rome Tor Vergata

Daniele Terlizzese
EIEF

SIdE webinar, July 1, 2020

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Introduction

- Knowing the prevalence of the COVID-19 infection in a population is very important for public health.
- But the fraction of people who test positive provides a biased picture...
- ...as the sample of those tested underrepresents the asymptomatic and paucisymptomatic.
- Several countries are carrying out large population surveys to ascertain the fraction of the population that has been infected by the virus.
- These surveys are costly, complicated to carry out, plagued by refusal to participate, and take time to complete.
- It is useful to come up with easy to implement – but perhaps less accurate – estimates of the prevalence of the infection based on readily available data.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Point vs. period prevalence

- Two distinct concepts of prevalence, with different meanings and usefulness:
  - point prevalence is the fraction of people who, at a given point in time, are infected (and therefore infectious);
  - period prevalence is the fraction of people who have been infected during a given period; in particular, the fraction of people who were ever infected.
- Point prevalence is more useful to monitor the risk of transmitting the infection, period prevalence is more useful for reporting purposes and to monitor the risk of being infected.
- A related paper by Manski and Molinari (2020) focuses on period prevalence.
- We focus instead on point prevalence.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Definitions and basic assumptions

- A person can be infected only once, is infectious as long as infected, and is no longer infectious when recovered.
- Time is measured in days. Population is constant.
- I_d: binary random variable equal to 1 if a person is infected on day d, and to 0 otherwise.
- We are interested in $\mathbb{P}(I_d = 1)$. A person who is infected at d may have become infected before d, and some of those who became infected before d may have recovered (or died) by d.
- T_d: binary random variable equal to 1 if a person has been first tested on day d, and to 0 otherwise.
- P_d: binary random variable equal to 1 if a person has first received a positive test result on day d, and to 0 otherwise.
- Assume that nobody is tested more than once, and that test results are immediately available. Under these assumptions, $P_d = 1 \Rightarrow T_d = 1$.
- We discuss later how to transform the data to approximate these assumptions.

Framework

- Drop the time suffix for simplicity.
- By the Law of Total Probability

$$\mathbb{P}(I = 1) = \mathbb{P}(I = 1|T = 1)\ \mathbb{P}(T = 1) + \mathbb{P}(I = 1|T = 0)\ \mathbb{P}(T = 0).$$

- We observe $\mathbb{P}(T = 1)$ and $\mathbb{P}(T = 0) = 1 - \mathbb{P}(T = 1)$, but $\mathbb{P}(I = 1|T = 1)$ and $\mathbb{P}(I = 1|T = 0)$ are unobserved.
- Thus, to estimate $\mathbb{P}(I = 1)$, we need information on both of them.
- First, consider the operational properties of viral tests.

Framework

- Drop the time suffix for simplicity.
- By the Law of Total Probability

$$\mathbb{P}(I = 1) = \mathbb{P}(I = 1|T = 1)\ \mathbb{P}(T = 1) + \mathbb{P}(I = 1|T = 0)\ \mathbb{P}(T = 0).$$

- We observe $\mathbb{P}(T = 1)$ and $\mathbb{P}(T = 0) = 1 - \mathbb{P}(T = 1)$, but $\mathbb{P}(I = 1|T = 1)$ and $\mathbb{P}(I = 1|T = 0)$ are unobserved.
- Thus, to estimate $\mathbb{P}(I = 1)$, we need information on both of them.
- First, consider the operational properties of viral tests.

Framework

- Drop the time suffix for simplicity.
- By the Law of Total Probability

$$\mathbb{P}(I = 1) = \mathbb{P}(I = 1|T = 1)\ \mathbb{P}(T = 1) + \mathbb{P}(I = 1|T = 0)\ \mathbb{P}(T = 0).$$

- We observe $\mathbb{P}(T = 1)$ and $\mathbb{P}(T = 0) = 1 - \mathbb{P}(T = 1)$, but $\mathbb{P}(I = 1|T = 1)$ and $\mathbb{P}(I = 1|T = 0)$ are unobserved.
- Thus, to estimate $\mathbb{P}(I = 1)$, we need information on both of them.
- First, consider the operational properties of viral tests.

Framework

- Drop the time suffix for simplicity.
- By the Law of Total Probability

$$\mathbb{P}(I = 1) = \mathbb{P}(I = 1|T = 1)\ \mathbb{P}(T = 1) + \mathbb{P}(I = 1|T = 0)\ \mathbb{P}(T = 0).$$

- We observe $\mathbb{P}(T = 1)$ and $\mathbb{P}(T = 0) = 1 - \mathbb{P}(T = 1)$, but $\mathbb{P}(I = 1|T = 1)$ and $\mathbb{P}(I = 1|T = 0)$ are unobserved.
- Thus, to estimate $\mathbb{P}(I = 1)$, we need information on both of them.
- First, consider the operational properties of viral tests.

Framework

- Drop the time suffix for simplicity.
- By the Law of Total Probability

$$\mathbb{P}(I = 1) = \mathbb{P}(I = 1|T = 1)\ \mathbb{P}(T = 1) + \mathbb{P}(I = 1|T = 0)\ \mathbb{P}(T = 0).$$

- We observe $\mathbb{P}(T = 1)$ and $\mathbb{P}(T = 0) = 1 - \mathbb{P}(T = 1)$, but $\mathbb{P}(I = 1|T = 1)$ and $\mathbb{P}(I = 1|T = 0)$ are unobserved.
- Thus, to estimate $\mathbb{P}(I = 1)$, we need information on both of them.
- First, consider the operational properties of viral tests.

Operational properties of viral tests

- Probability of false positives or Type-I error:

$$\mathbb{P}(P = 1 | I = 0, T = 1)$$

- There is a general consensus in the medical profession that the probability of Type-I error is negligible. Thus we assume

$$\mathbb{P}(P = 1 | I = 0, T = 1) = 0.$$

- Probability of false negatives or Type-II error:

$$\beta = \mathbb{P}(P = 0 | I = 1, T = 1).$$

- The general consensus is that β is nonnegligible and largely reflects issues with specimen collection (sample collected too early or too late, contaminated, or stored for too long).

- The available health literature suggests a range for β between .02 and .40, with a narrower range between .10 and .30 more often quoted.

Operational properties of viral tests

- Probability of false positives or Type-I error:

$$\mathbb{P}(P = 1 | I = 0, T = 1)$$

- There is a general consensus in the medical profession that the probability of Type-I error is negligible. Thus we assume

$$\mathbb{P}(P = 1 | I = 0, T = 1) = 0.$$

- Probability of false negatives or Type-II error:

$$\beta = \mathbb{P}(P = 0 | I = 1, T = 1).$$

- The general consensus is that β is nonnegligible and largely reflects issues with specimen collection (sample collected too early or too late, contaminated, or stored for too long).

- The available health literature suggests a range for β between .02 and .40, with a narrower range between .10 and .30 more often quoted.

Operational properties of viral tests

- Probability of false positives or Type-I error:

$$\mathbb{P}(P = 1 | I = 0, T = 1)$$

- There is a general consensus in the medical profession that the probability of Type-I error is negligible. Thus we assume

$$\mathbb{P}(P = 1 | I = 0, T = 1) = 0.$$

- Probability of false negatives or Type-II error:

$$\beta = \mathbb{P}(P = 0 | I = 1, T = 1).$$

- The general consensus is that β is nonnegligible and largely reflects issues with specimen collection (sample collected too early or too late, contaminated, or stored for too long).

- The available health literature suggests a range for β between .02 and .40, with a narrower range between .10 and .30 more often quoted.

Operational properties of viral tests

- Probability of false positives or Type-I error:

$$\mathbb{P}(P = 1 | I = 0, T = 1)$$

- There is a general consensus in the medical profession that the probability of Type-I error is negligible. Thus we assume

$$\mathbb{P}(P = 1 | I = 0, T = 1) = 0.$$

- Probability of false negatives or Type-II error:

$$\beta = \mathbb{P}(P = 0 | I = 1, T = 1).$$

- The general consensus is that β is nonnegligible and largely reflects issues with specimen collection (sample collected too early or too late, contaminated, or stored for too long).

- The available health literature suggests a range for β between .02 and .40, with a narrower range between .10 and .30 more often quoted.

Operational properties of viral tests

- Probability of false positives or Type-I error:

$$\mathbb{P}(P = 1 | I = 0, T = 1)$$

- There is a general consensus in the medical profession that the probability of Type-I error is negligible. Thus we assume

$$\mathbb{P}(P = 1 | I = 0, T = 1) = 0.$$

- Probability of false negatives or Type-II error:

$$\beta = \mathbb{P}(P = 0 | I = 1, T = 1).$$

- The general consensus is that β is nonnegligible and largely reflects issues with specimen collection (sample collected too early or too late, contaminated, or stored for too long).

- The available health literature suggests a range for β between .02 and .40, with a narrower range between .10 and .30 more often quoted.

Infection rate among the tested

- Under the assumption that $\mathbb{P}(P = 1|I = 0, T = 1) = 0$, we show that

$$\mathbb{P}(I = 1|T = 1) = \frac{\mathbb{P}(P = 1|T = 1)}{1 - \beta},$$

  where we used the fact that $P = 1 \Rightarrow T = 1$.

- This result expresses $\mathbb{P}(I = 1|T = 1)$ as a function of two quantities:
  - $\mathbb{P}(P = 1|T = 1)$ which is observed;
  - β for which we can make an educated guess based on medical knowledge.
- If $\beta = 0$, then $\mathbb{P}(I = 1|T = 1) = \mathbb{P}(P = 1|T = 1)$.

Infection rate among the tested

- Under the assumption that $\mathbb{P}(P = 1|I = 0, T = 1) = 0$, we show that

$$\mathbb{P}(I = 1|T = 1) = \frac{\mathbb{P}(P = 1|T = 1)}{1 - \beta},$$

  where we used the fact that $P = 1 \Rightarrow T = 1$.

- This result expresses $\mathbb{P}(I = 1|T = 1)$ as a function of two quantities:
  - $\mathbb{P}(P = 1|T = 1)$ which is observed;
  - β for which we can make an educated guess based on medical knowledge.
- If $\beta = 0$, then $\mathbb{P}(I = 1|T = 1) = \mathbb{P}(P = 1|T = 1)$.

Infection rate among the tested

- Under the assumption that $\mathbb{P}(P = 1|I = 0, T = 1) = 0$, we show that

$$\mathbb{P}(I = 1|T = 1) = \frac{\mathbb{P}(P = 1|T = 1)}{1 - \beta},$$

  where we used the fact that $P = 1 \Rightarrow T = 1$.

- This result expresses $\mathbb{P}(I = 1|T = 1)$ as a function of two quantities:
  - $\mathbb{P}(P = 1|T = 1)$ which is observed;
  - β for which we can make an educated guess based on medical knowledge.
- If $\beta = 0$, then $\mathbb{P}(I = 1|T = 1) = \mathbb{P}(P = 1|T = 1)$.

Infection rate among the tested

- Under the assumption that $\mathbb{P}(P = 1|I = 0, T = 1) = 0$, we show that

$$\mathbb{P}(I = 1|T = 1) = \frac{\mathbb{P}(P = 1|T = 1)}{1 - \beta},$$

  where we used the fact that $P = 1 \Rightarrow T = 1$.

- This result expresses $\mathbb{P}(I = 1|T = 1)$ as a function of two quantities:
  - $\mathbb{P}(P = 1|T = 1)$ which is observed;
  - β for which we can make an educated guess based on medical knowledge.
- If $\beta = 0$, then $\mathbb{P}(I = 1|T = 1) = \mathbb{P}(P = 1|T = 1)$.

Infection rate among the tested

- Under the assumption that $\mathbb{P}(P = 1|I = 0, T = 1) = 0$, we show that

$$\mathbb{P}(I = 1|T = 1) = \frac{\mathbb{P}(P = 1|T = 1)}{1 - \beta},$$

  where we used the fact that $P = 1 \Rightarrow T = 1$.

- This result expresses $\mathbb{P}(I = 1|T = 1)$ as a function of two quantities:
  - $\mathbb{P}(P = 1|T = 1)$ which is observed;
  - β for which we can make an educated guess based on medical knowledge.
- If $\beta = 0$, then $\mathbb{P}(I = 1|T = 1) = \mathbb{P}(P = 1|T = 1)$.

Infection rate among the untested

- Because swabs are mainly taken from subjects who have visible symptoms or are suspected of exposure to the infection, it is plausible to assume that

$$\mathbb{P}(I = 1|T = 0) \leq \mathbb{P}(I = 1|T = 1).$$

- If tests were random, the fraction of the infected would be the same among the tested and the untested.
- If the tested sample is biased towards those with higher infection risk, then $\mathbb{P}(I = 1|T = 0) < \mathbb{P}(I = 1|T = 1)$.
- Let $\mathbb{P}(I = 1|T = 1) > 0$ and define

$$\lambda = \frac{\mathbb{P}(I = 1|T = 0)}{\mathbb{P}(I = 1|T = 1)},$$

- Thus

$$\mathbb{P}(I = 1|T = 0) = \lambda\, \mathbb{P}(I = 1|T = 1),$$

with $0 < \lambda \leq 1$.

Infection rate among the untested

- Because swabs are mainly taken from subjects who have visible symptoms or are suspected of exposure to the infection, it is plausible to assume that

$$\mathbb{P}(I = 1|T = 0) \leq \mathbb{P}(I = 1|T = 1).$$

- If tests were random, the fraction of the infected would be the same among the tested and the untested.
- If the tested sample is biased towards those with higher infection risk, then $\mathbb{P}(I = 1|T = 0) < \mathbb{P}(I = 1|T = 1)$.
- Let $\mathbb{P}(I = 1|T = 1) > 0$ and define

$$\lambda = \frac{\mathbb{P}(I = 1|T = 0)}{\mathbb{P}(I = 1|T = 1)},$$

- Thus

$$\mathbb{P}(I = 1|T = 0) = \lambda\, \mathbb{P}(I = 1|T = 1),$$

with $0 < \lambda \leq 1$.

Infection rate among the untested

- Because swabs are mainly taken from subjects who have visible symptoms or are suspected of exposure to the infection, it is plausible to assume that

$$\mathbb{P}(I = 1|T = 0) \leq \mathbb{P}(I = 1|T = 1).$$

- If tests were random, the fraction of the infected would be the same among the tested and the untested.

- If the tested sample is biased towards those with higher infection risk, then $\mathbb{P}(I = 1|T = 0) < \mathbb{P}(I = 1|T = 1)$.

- Let $\mathbb{P}(I = 1|T = 1) > 0$ and define

$$\lambda = \frac{\mathbb{P}(I = 1|T = 0)}{\mathbb{P}(I = 1|T = 1)},$$

- Thus

$$\mathbb{P}(I = 1|T = 0) = \lambda\, \mathbb{P}(I = 1|T = 1),$$

with $0 < \lambda \leq 1$.

Infection rate among the untested

- Because swabs are mainly taken from subjects who have visible symptoms or are suspected of exposure to the infection, it is plausible to assume that

$$\mathbb{P}(I = 1|T = 0) \leq \mathbb{P}(I = 1|T = 1).$$

- If tests were random, the fraction of the infected would be the same among the tested and the untested.
- If the tested sample is biased towards those with higher infection risk, then $\mathbb{P}(I = 1|T = 0) < \mathbb{P}(I = 1|T = 1)$.
- Let $\mathbb{P}(I = 1|T = 1) > 0$ and define

$$\lambda = \frac{\mathbb{P}(I = 1|T = 0)}{\mathbb{P}(I = 1|T = 1)},$$

- Thus

$$\mathbb{P}(I = 1|T = 0) = \lambda\, \mathbb{P}(I = 1|T = 1),$$

with $0 < \lambda \leq 1$.

Infection rate among the untested

- Because swabs are mainly taken from subjects who have visible symptoms or are suspected of exposure to the infection, it is plausible to assume that

$$\mathbb{P}(I = 1|T = 0) \leq \mathbb{P}(I = 1|T = 1).$$

- If tests were random, the fraction of the infected would be the same among the tested and the untested.

- If the tested sample is biased towards those with higher infection risk, then $\mathbb{P}(I = 1|T = 0) < \mathbb{P}(I = 1|T = 1)$.

- Let $\mathbb{P}(I = 1|T = 1) > 0$ and define

$$\lambda = \frac{\mathbb{P}(I = 1|T = 0)}{\mathbb{P}(I = 1|T = 1)},$$

- Thus

$$\mathbb{P}(I = 1|T = 0) = \lambda\, \mathbb{P}(I = 1|T = 1),$$

with $0 < \lambda \leq 1$.

The estimating equation

- Putting all together, we have:

$$\begin{aligned}\mathbb{P}(I=1) &= \mathbb{P}(I=1|T=1)\ \mathbb{P}(T=1) + \mathbb{P}(I=1|T=0)\ \mathbb{P}(T=0) \\ &= \mathbb{P}(I=1|T=1)\ \mathbb{P}(T=1) + \lambda\,\mathbb{P}(I=1|T=1)\ \mathbb{P}(T=0) \\ &= \mathbb{P}(I=1|T=1)\,[\mathbb{P}(T=1) + \lambda\ \mathbb{P}(T=0)] \\ &= \frac{\mathbb{P}(P=1|T=1)}{1-\beta}\ [\mathbb{P}(T=1) + \lambda\ \mathbb{P}(T=0)]\,,\end{aligned}$$

  with $\mathbb{P}(T=0) = 1 - \mathbb{P}(T=1)$.

- This shows that $\mathbb{P}(I=1)$ is an increasing function of $\mathbb{P}(P=1|T=1)$ and $\mathbb{P}(T=1)$, which we can measure in the data, and an increasing function of β and λ, for which we have a plausible range of values.

The estimating equation

- Putting all together, we have:

$$\begin{aligned}
\mathbb{P}(I=1) &= \mathbb{P}(I=1|T=1)\ \mathbb{P}(T=1) + \mathbb{P}(I=1|T=0)\ \mathbb{P}(T=0) \\
&= \mathbb{P}(I=1|T=1)\ \mathbb{P}(T=1) + \lambda\,\mathbb{P}(I=1|T=1)\ \mathbb{P}(T=0) \\
&= \mathbb{P}(I=1|T=1)\,[\mathbb{P}(T=1) + \lambda\ \mathbb{P}(T=0)] \\
&= \frac{\mathbb{P}(P=1|T=1)}{1-\beta}\ [\mathbb{P}(T=1) + \lambda\ \mathbb{P}(T=0)]\,,
\end{aligned}$$

  with $\mathbb{P}(T=0) = 1 - \mathbb{P}(T=1)$.

- This shows that $\mathbb{P}(I=1)$ is an increasing function of $\mathbb{P}(P=1|T=1)$ and $\mathbb{P}(T=1)$, which we can measure in the data, and an increasing function of β and λ, for which we have a plausible range of values.

Can we say more about λ?

- Assume (wlog) that the test is perfect, but distinguish those who are symptomatic ($S = 1$) from those who are not ($S = 0$).
- Further assume that the infection rate among the asymptomatic, $\mathbb{P}(I = 1|S = 0)$, is positive and smaller than among the symptomatic, that is:

$$\mu = \frac{\mathbb{P}(I = 1|S = 1)}{\mathbb{P}(I = 1|S = 0)} \geq 1.$$

- Let $n = \mathbb{P}(T = 1)$ be the fraction of the population that is tested, $\pi = \mathbb{P}(S = 1)$ the fraction of the population that is symptomatic, and $p = \mathbb{P}(S = 1|T = 1)$ the fraction of the symptomatic among the tested.
- As a measure of the bias implicit in the testing protocol, define

$$\gamma = \frac{p}{\pi}.$$

- If the sample is biased towards the symptomatic, then $p > \pi$, that is, $\gamma > 1$.

Can we say more about λ?

- Assume (wlog) that the test is perfect, but distinguish those who are symptomatic ($S = 1$) from those who are not ($S = 0$).
- Further assume that the infection rate among the asymptomatic, $\mathbb{P}(I = 1|S = 0)$, is positive and smaller than among the symptomatic, that is:

$$\mu = \frac{\mathbb{P}(I = 1|S = 1)}{\mathbb{P}(I = 1|S = 0)} \geq 1.$$

- Let $n = \mathbb{P}(T = 1)$ be the fraction of the population that is tested, $\pi = \mathbb{P}(S = 1)$ the fraction of the population that is symptomatic, and $p = \mathbb{P}(S = 1|T = 1)$ the fraction of the symptomatic among the tested.
- As a measure of the bias implicit in the testing protocol, define

$$\gamma = \frac{p}{\pi}.$$

- If the sample is biased towards the symptomatic, then $p > \pi$, that is, $\gamma > 1$.

Can we say more about λ?

- Assume (wlog) that the test is perfect, but distinguish those who are symptomatic ($S = 1$) from those who are not ($S = 0$).
- Further assume that the infection rate among the asymptomatic, $\mathbb{P}(I = 1|S = 0)$, is positive and smaller than among the symptomatic, that is:

$$\mu = \frac{\mathbb{P}(I = 1|S = 1)}{\mathbb{P}(I = 1|S = 0)} \geq 1.$$

- Let $n = \mathbb{P}(T = 1)$ be the fraction of the population that is tested, $\pi = \mathbb{P}(S = 1)$ the fraction of the population that is symptomatic, and $p = \mathbb{P}(S = 1|T = 1)$ the fraction of the symptomatic among the tested.
- As a measure of the bias implicit in the testing protocol, define

$$\gamma = \frac{p}{\pi}.$$

- If the sample is biased towards the symptomatic, then $p > \pi$, that is, $\gamma > 1$.

Can we say more about λ?

- Assume (wlog) that the test is perfect, but distinguish those who are symptomatic ($S = 1$) from those who are not ($S = 0$).
- Further assume that the infection rate among the asymptomatic, $\mathbb{P}(I = 1|S = 0)$, is positive and smaller than among the symptomatic, that is:

$$\mu = \frac{\mathbb{P}(I = 1|S = 1)}{\mathbb{P}(I = 1|S = 0)} \geq 1.$$

- Let $n = \mathbb{P}(T = 1)$ be the fraction of the population that is tested, $\pi = \mathbb{P}(S = 1)$ the fraction of the population that is symptomatic, and $p = \mathbb{P}(S = 1|T = 1)$ the fraction of the symptomatic among the tested.
- As a measure of the bias implicit in the testing protocol, define

$$\gamma = \frac{p}{\pi}.$$

- If the sample is biased towards the symptomatic, then $p > \pi$, that is, $\gamma > 1$.

Can we say more about λ?

- Assume (wlog) that the test is perfect, but distinguish those who are symptomatic ($S = 1$) from those who are not ($S = 0$).
- Further assume that the infection rate among the asymptomatic, $\mathbb{P}(I = 1|S = 0)$, is positive and smaller than among the symptomatic, that is:

$$\mu = \frac{\mathbb{P}(I = 1|S = 1)}{\mathbb{P}(I = 1|S = 0)} \geq 1.$$

- Let $n = \mathbb{P}(T = 1)$ be the fraction of the population that is tested, $\pi = \mathbb{P}(S = 1)$ the fraction of the population that is symptomatic, and $p = \mathbb{P}(S = 1|T = 1)$ the fraction of the symptomatic among the tested.
- As a measure of the bias implicit in the testing protocol, define

$$\gamma = \frac{p}{\pi}.$$

- If the sample is biased towards the symptomatic, then $p > \pi$, that is, $\gamma > 1$.

The structure of λ

- Among the tested cases, the probability of being infected is:

$$\mathbb{P}(I=1|T=1) = \mathbb{P}(I=1|S=1,T=1)\,p + \mathbb{P}(I=1|S=0,T=1)(1-p).$$

- Our key assumption is that, conditioning on symptomatology, the probability of being infected is the same in the tested sample and the population:

$$\mathbb{P}(I=1|S=s,T=1) = \mathbb{P}(I=1|S=s), \qquad s=0,1.$$

Under this assumption,

$$\mathbb{P}(I=1|T=1) = [(\mu-1)p+1]\ \mathbb{P}(I=1|S=0).$$

- Among the untested cases, the probability of being infected can be shown to be:

$$\mathbb{P}(I=1|T=0) = \frac{(\mu-1)(\pi-np)+1-n}{1-n}\ \mathbb{P}(I=1|S=0).$$

- Combining these two results, gives:

$$\lambda = \frac{\mathbb{P}(I=1|T=0)}{\mathbb{P}(I=1|T=1)} = \frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}.$$

- Information on n, γ, μ, or π may help narrow the range of values for λ and therefore for $\mathbb{P}(I=1)$. In particular, $\lambda = 1$ when $\gamma = 1$.

The structure of λ

- Among the tested cases, the probability of being infected is:

$$\mathbb{P}(I=1|T=1) = \mathbb{P}(I=1|S=1,T=1)\,p + \mathbb{P}(I=1|S=0,T=1)(1-p).$$

- Our key assumption is that, conditioning on symptomatology, the probability of being infected is the same in the tested sample and the population:

$$\mathbb{P}(I=1|S=s,T=1) = \mathbb{P}(I=1|S=s), \qquad s=0,1.$$

Under this assumption,

$$\mathbb{P}(I=1|T=1) = [(\mu-1)p+1]\ \mathbb{P}(I=1|S=0).$$

- Among the untested cases, the probability of being infected can be shown to be:

$$\mathbb{P}(I=1|T=0) = \frac{(\mu-1)(\pi-np)+1-n}{1-n}\ \mathbb{P}(I=1|S=0).$$

- Combining these two results, gives:

$$\lambda = \frac{\mathbb{P}(I=1|T=0)}{\mathbb{P}(I=1|T=1)} = \frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}.$$

- Information on n, γ, μ, or π may help narrow the range of values for λ and therefore for $\mathbb{P}(I=1)$. In particular, $\lambda = 1$ when $\gamma = 1$.

The structure of λ

- Among the tested cases, the probability of being infected is:

$$\mathbb{P}(I=1|T=1) = \mathbb{P}(I=1|S=1,T=1)\,p + \mathbb{P}(I=1|S=0,T=1)(1-p).$$

- Our key assumption is that, conditioning on symptomatology, the probability of being infected is the same in the tested sample and the population:

$$\mathbb{P}(I=1|S=s,T=1) = \mathbb{P}(I=1|S=s), \qquad s=0,1.$$

Under this assumption,

$$\mathbb{P}(I=1|T=1) = [(\mu-1)p+1]\ \mathbb{P}(I=1|S=0).$$

- Among the untested cases, the probability of being infected can be shown to be:

$$\mathbb{P}(I=1|T=0) = \frac{(\mu-1)(\pi-np)+1-n}{1-n}\ \mathbb{P}(I=1|S=0).$$

- Combining these two results, gives:

$$\lambda = \frac{\mathbb{P}(I=1|T=0)}{\mathbb{P}(I=1|T=1)} = \frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}.$$

- Information on n, γ, μ, or π may help narrow the range of values for λ and therefore for $\mathbb{P}(I=1)$. In particular, $\lambda = 1$ when $\gamma = 1$.

The structure of λ

- Among the tested cases, the probability of being infected is:

$$\mathbb{P}(I=1|T=1) = \mathbb{P}(I=1|S=1,T=1)\,p + \mathbb{P}(I=1|S=0,T=1)(1-p).$$

- Our key assumption is that, conditioning on symptomatology, the probability of being infected is the same in the tested sample and the population:

$$\mathbb{P}(I=1|S=s,T=1) = \mathbb{P}(I=1|S=s), \qquad s=0,1.$$

Under this assumption,

$$\mathbb{P}(I=1|T=1) = [(\mu-1)p+1]\ \mathbb{P}(I=1|S=0).$$

- Among the untested cases, the probability of being infected can be shown to be:

$$\mathbb{P}(I=1|T=0) = \frac{(\mu-1)(\pi-np)+1-n}{1-n}\ \mathbb{P}(I=1|S=0).$$

- Combining these two results, gives:

$$\lambda = \frac{\mathbb{P}(I=1|T=0)}{\mathbb{P}(I=1|T=1)} = \frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}.$$

- Information on n, γ, μ, or π may help narrow the range of values for λ and therefore for $\mathbb{P}(I=1)$. In particular, $\lambda = 1$ when $\gamma = 1$.

The structure of λ

- Among the tested cases, the probability of being infected is:

$$\mathbb{P}(I=1|T=1) = \mathbb{P}(I=1|S=1,T=1)\,p + \mathbb{P}(I=1|S=0,T=1)(1-p).$$

- Our key assumption is that, conditioning on symptomatology, the probability of being infected is the same in the tested sample and the population:

$$\mathbb{P}(I=1|S=s,T=1) = \mathbb{P}(I=1|S=s), \qquad s=0,1.$$

  Under this assumption,

$$\mathbb{P}(I=1|T=1) = [(\mu-1)p+1]\;\mathbb{P}(I=1|S=0).$$

- Among the untested cases, the probability of being infected can be shown to be:

$$\mathbb{P}(I=1|T=0) = \frac{(\mu-1)(\pi-np)+1-n}{1-n}\;\mathbb{P}(I=1|S=0).$$

- Combining these two results, gives:

$$\lambda = \frac{\mathbb{P}(I=1|T=0)}{\mathbb{P}(I=1|T=1)} = \frac{(\mu-1)(1-n\gamma)\pi+1-n}{(1-n)[\gamma\pi(\mu-1)+1]}.$$

- Information on n, γ, μ, or π may help narrow the range of values for λ and therefore for $\mathbb{P}(I=1)$. In particular, $\lambda = 1$ when $\gamma = 1$.

Calibrating λ

- As an example, consider the data from Vo', the small town with the first COVID-19 death in Italy, that conducted an almost complete testing of its population between late February and early March, 2020.
- From these data we can set $\pi = .067$ and $\mu = 17.9$.
- We also set $n = .0004$ (the testing rate in Italy on June 20, 2020).
- Given these values, the next slide shows how λ would vary with the bias γ in the testing protocol when $\mu = 17.9$.
- Specifically, λ would vary between .11 (when $\gamma = 1/.067 = 15$) and .65 (when $\gamma = 2$, i.e., the sample fraction of the symptomatic is twice the population fraction).

Calibrating λ

- As an example, consider the data from Vo', the small town with the first COVID-19 death in Italy, that conducted an almost complete testing of its population between late February and early March, 2020.
- From these data we can set $\pi = .067$ and $\mu = 17.9$.
- We also set $n = .0004$ (the testing rate in Italy on June 20, 2020).
- Given these values, the next slide shows how λ would vary with the bias γ in the testing protocol when $\mu = 17.9$.
- Specifically, λ would vary between .11 (when $\gamma = 1/.067 = 15$) and .65 (when $\gamma = 2$, i.e., the sample fraction of the symptomatic is twice the population fraction).

Calibrating λ

- As an example, consider the data from Vo', the small town with the first COVID-19 death in Italy, that conducted an almost complete testing of its population between late February and early March, 2020.
- From these data we can set $\pi = .067$ and $\mu = 17.9$.
- We also set $n = .0004$ (the testing rate in Italy on June 20, 2020).
- Given these values, the next slide shows how λ would vary with the bias γ in the testing protocol when $\mu = 17.9$.
- Specifically, λ would vary between .11 (when $\gamma = 1/.067 = 15$) and .65 (when $\gamma = 2$, i.e., the sample fraction of the symptomatic is twice the population fraction).

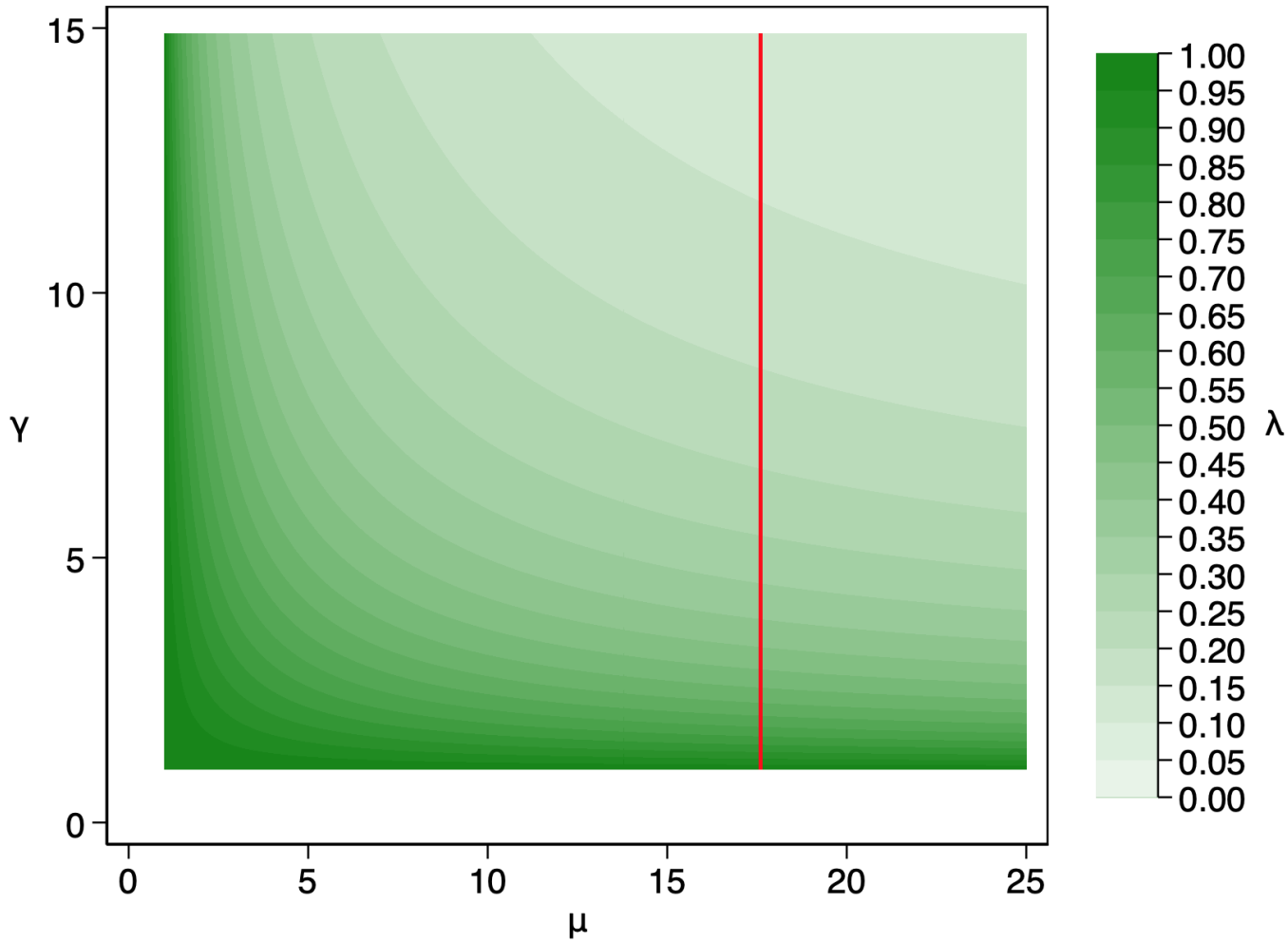
Calibrating λ

- As an example, consider the data from Vo', the small town with the first COVID-19 death in Italy, that conducted an almost complete testing of its population between late February and early March, 2020.
- From these data we can set $\pi = .067$ and $\mu = 17.9$.
- We also set $n = .0004$ (the testing rate in Italy on June 20, 2020).
- Given these values, the next slide shows how λ would vary with the bias γ in the testing protocol when $\mu = 17.9$.
- Specifically, λ would vary between .11 (when $\gamma = 1/.067 = 15$) and .65 (when $\gamma = 2$, i.e., the sample fraction of the symptomatic is twice the population fraction).

Calibrating λ

- As an example, consider the data from Vo', the small town with the first COVID-19 death in Italy, that conducted an almost complete testing of its population between late February and early March, 2020.
- From these data we can set $\pi = .067$ and $\mu = 17.9$.
- We also set $n = .0004$ (the testing rate in Italy on June 20, 2020).
- Given these values, the next slide shows how λ would vary with the bias γ in the testing protocol when $\mu = 17.9$.
- Specifically, λ would vary between .11 (when $\gamma = 1/.067 = 15$) and .65 (when $\gamma = 2$, i.e., the sample fraction of the symptomatic is twice the population fraction).

Contour plot of λ as a function of γ and μ



Detour on point vs. period prevalence

- Both concepts are useful. Can both be recovered from the available data?
- Assume that the population is of constant size N, the infection lasts at most 2 periods, nobody dies, a fraction δ of the newly infected recovers after 1 period and the remaining $1-\delta$ after 2 period.
- Hence, in any day $d \geq 1$, the number of the currently infected is

$$I_d = I_d^* + (1-\delta) I_{d-1}^*,$$

  where I_d^* are the newly infected on day d and $I_0^* = 0$.
- Point and period prevalence at time d in the population are, respectively,

$$\alpha_d = \frac{I_d}{N} = \frac{I_d^* + (1-\delta) I_{d-1}^*}{N}, \qquad \eta_d = \frac{\sum_{i=1}^{d} I_i^*}{N}.$$

- While α_d is obtained from the currently infected (who might have been first infected in previous periods), η_d is constructed by adding-up all the newly infected up to day d.

Detour on point vs. period prevalence

- Both concepts are useful. Can both be recovered from the available data?
- Assume that the population is of constant size N, the infection lasts at most 2 periods, nobody dies, a fraction δ of the newly infected recovers after 1 period and the remaining $1-\delta$ after 2 period.
- Hence, in any day $d \geq 1$, the number of the currently infected is

$$I_d = I_d^* + (1-\delta) I_{d-1}^*,$$

where I_d^* are the newly infected on day d and $I_0^* = 0$.

- Point and period prevalence at time d in the population are, respectively,

$$\alpha_d = \frac{I_d}{N} = \frac{I_d^* + (1-\delta) I_{d-1}^*}{N}, \qquad \eta_d = \frac{\sum_{i=1}^{d} I_i^*}{N}.$$

- While α_d is obtained from the currently infected (who might have been first infected in previous periods), η_d is constructed by adding-up all the newly infected up to day d.

Detour on point vs. period prevalence

- Both concepts are useful. Can both be recovered from the available data?
- Assume that the population is of constant size N, the infection lasts at most 2 periods, nobody dies, a fraction δ of the newly infected recovers after 1 period and the remaining $1-\delta$ after 2 period.
- Hence, in any day $d \geq 1$, the number of the currently infected is

$$I_d = I_d^* + (1-\delta) I_{d-1}^*,$$

where I_d^* are the newly infected on day d and $I_0^* = 0$.

- Point and period prevalence at time d in the population are, respectively,

$$\alpha_d = \frac{I_d}{N} = \frac{I_d^* + (1-\delta) I_{d-1}^*}{N}, \qquad \eta_d = \frac{\sum_{i=1}^{d} I_i^*}{N}.$$

- While α_d is obtained from the currently infected (who might have been first infected in previous periods), η_d is constructed by adding-up all the newly infected up to day d.

Detour on point vs. period prevalence

- Both concepts are useful. Can both be recovered from the available data?
- Assume that the population is of constant size N, the infection lasts at most 2 periods, nobody dies, a fraction δ of the newly infected recovers after 1 period and the remaining $1-\delta$ after 2 period.
- Hence, in any day $d \geq 1$, the number of the currently infected is

$$I_d = I_d^* + (1-\delta) I_{d-1}^*,$$

  where I_d^* are the newly infected on day d and $I_0^* = 0$.
- Point and period prevalence at time d in the population are, respectively,

$$\alpha_d = \frac{I_d}{N} = \frac{I_d^* + (1-\delta) I_{d-1}^*}{N}, \qquad \eta_d = \frac{\sum_{i=1}^{d} I_i^*}{N}.$$

- While α_d is obtained from the currently infected (who might have been first infected in previous periods), η_d is constructed by adding-up all the newly infected up to day d.

Detour on point vs. period prevalence

- Both concepts are useful. Can both be recovered from the available data?
- Assume that the population is of constant size N, the infection lasts at most 2 periods, nobody dies, a fraction δ of the newly infected recovers after 1 period and the remaining $1-\delta$ after 2 period.
- Hence, in any day $d \geq 1$, the number of the currently infected is

$$I_d = I_d^* + (1-\delta) I_{d-1}^*,$$

  where I_d^* are the newly infected on day d and $I_0^* = 0$.
- Point and period prevalence at time d in the population are, respectively,

$$\alpha_d = \frac{I_d}{N} = \frac{I_d^* + (1-\delta) I_{d-1}^*}{N}, \qquad \eta_d = \frac{\sum_{i=1}^{d} I_i^*}{N}.$$

- While α_d is obtained from the currently infected (who might have been first infected in previous periods), η_d is constructed by adding-up all the newly infected up to day d.

Point vs. period prevalence (cntd.)

- The number of people who just recovered in period d is

$$R_d = \delta I^*_{d-1} + (1-\delta) I^*_{d-2},$$

with $R_1 = 0$ and $R_2 = \delta I^*_1$, while the total number of people who recovered from past infections is

$$RT_d = \sum_{i=1}^{d} R_i.$$

- It can be shown that

$$\sum_{i=1}^{d} I^*_i = I_d + \sum_{i=1}^{d} R_i = I_d + RT_d.$$

Therefore

$$\eta_d = \alpha_d + \frac{RT_d}{N}.$$

- This is the link between the two concepts in the population: starting from point prevalence, to obtain period prevalence we need data on current and past recoveries.

Point vs. period prevalence (cntd.)

- The number of people who just recovered in period d is

$$R_d = \delta I^*_{d-1} + (1-\delta) I^*_{d-2},$$

with $R_1 = 0$ and $R_2 = \delta I^*_1$, while the total number of people who recovered from past infections is

$$RT_d = \sum_{i=1}^{d} R_i.$$

- It can be shown that

$$\sum_{i=1}^{d} I^*_i = I_d + \sum_{i=1}^{d} R_i = I_d + RT_d.$$

Therefore

$$\eta_d = \alpha_d + \frac{RT_d}{N}.$$

- This is the link between the two concepts in the population: starting from point prevalence, to obtain period prevalence we need data on current and past recoveries.

Point vs. period prevalence (cntd.)

- The number of people who just recovered in period d is

$$R_d = \delta I^*_{d-1} + (1-\delta) I^*_{d-2},$$

with $R_1 = 0$ and $R_2 = \delta I^*_1$, while the total number of people who recovered from past infections is

$$RT_d = \sum_{i=1}^{d} R_i.$$

- It can be shown that

$$\sum_{i=1}^{d} I^*_i = I_d + \sum_{i=1}^{d} R_i = I_d + RT_d.$$

Therefore

$$\eta_d = \alpha_d + \frac{RT_d}{N}.$$

- This is the link between the two concepts in the population: starting from point prevalence, to obtain period prevalence we need data on current and past recoveries.

From population to samples

- We do not observe the population, but only a sample drawn from it.
- Consider a simple example in which the virological test is perfect and sampling is at random.
- None of these assumptions is realistic, but they provide a tractable and useful benchmark.
- This example is also of practical interest as some countries are embarking on random sampling of the general population.

From population to samples

- We do not observe the population, but only a sample drawn from it.
- Consider a simple example in which the virological test is perfect and sampling is at random.
- None of these assumptions is realistic, but they provide a tractable and useful benchmark.
- This example is also of practical interest as some countries are embarking on random sampling of the general population.

From population to samples

- We do not observe the population, but only a sample drawn from it.
- Consider a simple example in which the virological test is perfect and sampling is at random.
- None of these assumptions is realistic, but they provide a tractable and useful benchmark.
- This example is also of practical interest as some countries are embarking on random sampling of the general population.

From population to samples

- We do not observe the population, but only a sample drawn from it.
- Consider a simple example in which the virological test is perfect and sampling is at random.
- None of these assumptions is realistic, but they provide a tractable and useful benchmark.
- This example is also of practical interest as some countries are embarking on random sampling of the general population.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Example

- On any day d, a random sample of size n_d is tested. Let P_d be the number of people who test positive in the sample. Since $\mathbb{E}[P_d] = n_d \alpha_d$, an unbiased estimate of point prevalence is $\widehat{\alpha}_d = P_d / n_d$.
- To estimate period prevalence we need data on the number of people in the sample who recovered from past infections.
- On day d, they are (on average) $n_d \, RT_d / N$. Dividing by n_d gives an unbiased estimate of RT_d / N; adding $\widehat{\alpha}_d$, gives an unbiased estimate of period prevalence.
- Notice however that, to identify the recovered, the subjects in the sample should be tested with a serological test to detect past infections.
- This is not how the available data are constructed.
- The available data only keep track of future recoveries among the currently infected, but we do not know whether those who test negative at a point in time had been infected in the past and are now recovered.
- Having a biased sample would only add a further layer of difficulty.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Data

- Daily time series collected by the Dipartimento della Protezione Civile (DPC) from the COVID-19 surveillance system on:
  - total number of confirmed cases ("totale casi") and number of confirmed new cases ("nuovi positivi"), from February 24, 2020;
  - total number of swabs ("tamponi"), from February 24, 2020;
  - total number of tested cases ("casi testati"), from April 19, 2020.
- The most recent date is July 1, 2020.
- Case definitions follow the criteria established by the ECDPC, which in turn follow those published and updated by the WHO.
- Population data from Istat: Resident population as of January 1, 2019.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Remarks

- The series of new confirmed cases gives the number of subjects who first test positive on day d, i.e., the number of subjects for whom $P_d = 1$. They may have become infected on a previous date.
- The number of swabs overstates the number of people actually tested because many subjects are tested repeatedly.
- The series of tested cases records the total number of subjects from whom a swab was taken, thus eliminating the duplications contained in the swabs series.
- Besides duplications, the timing between swabs and tests is not fully aligned, as the number of tests results obtained on day d include results from swabs taken before d and excludes swabs taken in d for which results are not yet available.
- To take care of this (and to remove day-of-week effects), we use 7-day moving averages.
- The series of tested cases in one of the regions (Lazio) has a clear break on April 24, 2020, most likely because of an initial reporting error. We therefore consider the data starting from April 25, 2020.

Estimation

- To estimate $\mathbb{P}(I = 1)$ we first compute $\mathbb{P}(P = 1|T = 1)$ as the ratio between the centered 7-day moving average, or MA(7), of the confirmed new cases and the MA(7) of the difference in tested cases.
- We also compute $\mathbb{P}(T = 1)$ as the ratio between the MA(7) of the difference in tested cases and the resident population.
- As already mentioned, medical expertise suggests a relatively narrow range for β, say between .10 and .30. To be conservative we consider the wider range from .01 to .50, but note that a test with β as large as .50 is virtually useless (and dangerous!).
- As to λ we conservatively allow for a wide range, from .01 to .99. The data from Vo' would suggest a narrower range, between .10 and .65.

Estimation

- To estimate $\mathbb{P}(I = 1)$ we first compute $\mathbb{P}(P = 1|T = 1)$ as the ratio between the centered 7-day moving average, or MA(7), of the confirmed new cases and the MA(7) of the difference in tested cases.
- We also compute $\mathbb{P}(T = 1)$ as the ratio between the MA(7) of the difference in tested cases and the resident population.
- As already mentioned, medical expertise suggests a relatively narrow range for β, say between .10 and .30. To be conservative we consider the wider range from .01 to .50, but note that a test with β as large as .50 is virtually useless (and dangerous!).
- As to λ we conservatively allow for a wide range, from .01 to .99. The data from Vo' would suggest a narrower range, between .10 and .65.

Estimation

- To estimate $\mathbb{P}(I = 1)$ we first compute $\mathbb{P}(P = 1|T = 1)$ as the ratio between the centered 7-day moving average, or MA(7), of the confirmed new cases and the MA(7) of the difference in tested cases.
- We also compute $\mathbb{P}(T = 1)$ as the ratio between the MA(7) of the difference in tested cases and the resident population.
- As already mentioned, medical expertise suggests a relatively narrow range for β, say between .10 and .30. To be conservative we consider the wider range from .01 to .50, but note that a test with β as large as .50 is virtually useless (and dangerous!).
- As to λ we conservatively allow for a wide range, from .01 to .99. The data from Vo' would suggest a narrower range, between .10 and .65.

Estimation

- To estimate $\mathbb{P}(I = 1)$ we first compute $\mathbb{P}(P = 1|T = 1)$ as the ratio between the centered 7-day moving average, or MA(7), of the confirmed new cases and the MA(7) of the difference in tested cases.
- We also compute $\mathbb{P}(T = 1)$ as the ratio between the MA(7) of the difference in tested cases and the resident population.
- As already mentioned, medical expertise suggests a relatively narrow range for β, say between .10 and .30. To be conservative we consider the wider range from .01 to .50, but note that a test with β as large as .50 is virtually useless (and dangerous!).
- As to λ we conservatively allow for a wide range, from .01 to .99. The data from Vo' would suggest a narrower range, between .10 and .65.

Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020

Region	$\mathbb{P}(T = 1)$	$\mathbb{P}(P = 1\|T = 1)$
Abruzzo	.0005	.0007
Basilicata	.0005	.0000
Calabria	.0004	.0022
Campania	.0001	.0041
Emilia-Romagna	.0008	.0063
Friuli Venezia Giulia	.0008	.0011
Lazio	.0004	.0038
Liguria	.0004	.0106
Lombardia	.0006	.0273
Marche	.0004	.0027
Molise	.0007	.0035
Piemonte	.0004	.0151
Puglia	.0003	.0014
Sardegna	.0005	.0007
Sicilia	.0003	.0008
Toscana	.0004	.0026
Trentino-Alto Adige	.0008	.0069
Umbria	.0005	.0003
Valle d'Aosta	.0006	.0035
Veneto	.0005	.0018
Italy	.0004	.0078

Remarks

- Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020.
- While $\mathbb{P}(T = 1)$ does not vary much across regions, differences in $\mathbb{P}(P = 1|T = 1)$ are large reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing.
- $\mathbb{P}(T = 1)$ highest in Emilia-Romagna, Friuli Venezia Giulia, and Trentino-Alto Adige, lowest in Campania.
- $\mathbb{P}(P = 1|T = 1)$ highest in the North-West (Liguria, Piedmont, and especially Lombardy), lowest in Basilicata and Umbria.
- Very weak positive correlation between $\mathbb{P}(P = 1|T = 1)$ and $\mathbb{P}(T = 1)$.

Remarks

- Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020.
- While $\mathbb{P}(T = 1)$ does not vary much across regions, differences in $\mathbb{P}(P = 1|T = 1)$ are large reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing.
- $\mathbb{P}(T = 1)$ highest in Emilia-Romagna, Friuli Venezia Giulia, and Trentino-Alto Adige, lowest in Campania.
- $\mathbb{P}(P = 1|T = 1)$ highest in the North-West (Liguria, Piedmont, and especially Lombardy), lowest in Basilicata and Umbria.
- Very weak positive correlation between $\mathbb{P}(P = 1|T = 1)$ and $\mathbb{P}(T = 1)$.

Remarks

- Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020.
- While $\mathbb{P}(T = 1)$ does not vary much across regions, differences in $\mathbb{P}(P = 1|T = 1)$ are large reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing.
- $\mathbb{P}(T = 1)$ highest in Emilia-Romagna, Friuli Venezia Giulia, and Trentino-Alto Adige, lowest in Campania.
- $\mathbb{P}(P = 1|T = 1)$ highest in the North-West (Liguria, Piedmont, and especially Lombardy), lowest in Basilicata and Umbria.
- Very weak positive correlation between $\mathbb{P}(P = 1|T = 1)$ and $\mathbb{P}(T = 1)$.

Remarks

- Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020.
- While $\mathbb{P}(T = 1)$ does not vary much across regions, differences in $\mathbb{P}(P = 1|T = 1)$ are large reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing.
- $\mathbb{P}(T = 1)$ highest in Emilia-Romagna, Friuli Venezia Giulia, and Trentino-Alto Adige, lowest in Campania.
- $\mathbb{P}(P = 1|T = 1)$ highest in the North-West (Liguria, Piedmont, and especially Lombardy), lowest in Basilicata and Umbria.
- Very weak positive correlation between $\mathbb{P}(P = 1|T = 1)$ and $\mathbb{P}(T = 1)$.

Remarks

- Table 1: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$, June 20, 2020.
- While $\mathbb{P}(T = 1)$ does not vary much across regions, differences in $\mathbb{P}(P = 1|T = 1)$ are large reflecting substantial regional differences in both the intensity of the epidemic and the bias in testing.
- $\mathbb{P}(T = 1)$ highest in Emilia-Romagna, Friuli Venezia Giulia, and Trentino-Alto Adige, lowest in Campania.
- $\mathbb{P}(P = 1|T = 1)$ highest in the North-West (Liguria, Piedmont, and especially Lombardy), lowest in Basilicata and Umbria.
- Very weak positive correlation between $\mathbb{P}(P = 1|T = 1)$ and $\mathbb{P}(T = 1)$.

Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β)

	β										
λ	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
.01	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.05	.000	.000	.000	.000	.000	.001	.001	.001	.001	.001	.001
.10	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001	.002
.15	.001	.001	.001	.001	.001	.002	.002	.002	.002	.002	.002
.20	.002	.002	.002	.002	.002	.002	.002	.002	.003	.003	.003
.25	.002	.002	.002	.002	.002	.003	.003	.003	.003	.004	.004
.30	.002	.002	.003	.003	.003	.003	.003	.004	.004	.004	.005
.35	.003	.003	.003	.003	.003	.004	.004	.004	.005	.005	.005
.40	.003	.003	.003	.004	.004	.004	.004	.005	.005	.006	.006
.45	.004	.004	.004	.004	.004	.005	.005	.005	.006	.006	.007
.50	.004	.004	.004	.005	.005	.005	.006	.006	.007	.007	.008
.55	.004	.005	.005	.005	.005	.006	.006	.007	.007	.008	.009
.60	.005	.005	.005	.006	.006	.006	.007	.007	.008	.009	.009
.65	.005	.005	.006	.006	.006	.007	.007	.008	.008	.009	.010
.70	.006	.006	.006	.006	.007	.007	.008	.008	.009	.010	.011
.75	.006	.006	.007	.007	.007	.008	.008	.009	.010	.011	.012
.80	.006	.007	.007	.007	.008	.008	.009	.010	.010	.011	.013
.85	.007	.007	.007	.008	.008	.009	.010	.010	.011	.012	.013
.90	.007	.007	.008	.008	.009	.009	.010	.011	.012	.013	.014
.95	.008	.008	.008	.009	.009	.010	.011	.011	.012	.014	.015
.99	.008	.008	.009	.009	.010	.010	.011	.012	.013	.014	.015

Remarks

- Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β), June 20, 2020.
- $\mathbb{P}(I = 1)$ increases with both λ and β, ranging between 0 for $(\lambda, \beta) = (0, 0)$ and .015 for $(\lambda, \beta) = (.99, .50)$.
- Given the population size of 60.4 million, the upper bound corresponds to about 900 thousands infected people as of June 20, 2020.
- In red, estimated $\mathbb{P}(I = 1)$ for the restricted range $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.
- Estimated $\mathbb{P}(I = 1)$ ranges between .001 when $(\lambda, \beta) = (.10, .10)$ and .007 when $(\lambda, \beta) = (.65, .30)$, corresponding to a range between 60 and 423 thousands infected people as of June 20, 2020.

Remarks

- Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β), June 20, 2020.
- $\mathbb{P}(I = 1)$ increases with both λ and β, ranging between 0 for $(\lambda, \beta) = (0, 0)$ and .015 for $(\lambda, \beta) = (.99, .50)$.
- Given the population size of 60.4 million, the upper bound corresponds to about 900 thousands infected people as of June 20, 2020.
- In red, estimated $\mathbb{P}(I = 1)$ for the restricted range $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.
- Estimated $\mathbb{P}(I = 1)$ ranges between .001 when $(\lambda, \beta) = (.10, .10)$ and .007 when $(\lambda, \beta) = (.65, .30)$, corresponding to a range between 60 and 423 thousands infected people as of June 20, 2020.

Remarks

- Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β), June 20, 2020.
- $\mathbb{P}(I = 1)$ increases with both λ and β, ranging between 0 for $(\lambda, \beta) = (0, 0)$ and .015 for $(\lambda, \beta) = (.99, .50)$.
- Given the population size of 60.4 million, the upper bound corresponds to about 900 thousands infected people as of June 20, 2020.
- In red, estimated $\mathbb{P}(I = 1)$ for the restricted range $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.
- Estimated $\mathbb{P}(I = 1)$ ranges between .001 when $(\lambda, \beta) = (.10, .10)$ and .007 when $(\lambda, \beta) = (.65, .30)$, corresponding to a range between 60 and 423 thousands infected people as of June 20, 2020.

Remarks

- Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β), June 20, 2020.
- $\mathbb{P}(I = 1)$ increases with both λ and β, ranging between 0 for $(\lambda, \beta) = (0, 0)$ and .015 for $(\lambda, \beta) = (.99, .50)$.
- Given the population size of 60.4 million, the upper bound corresponds to about 900 thousands infected people as of June 20, 2020.
- In red, estimated $\mathbb{P}(I = 1)$ for the restricted range $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.
- Estimated $\mathbb{P}(I = 1)$ ranges between .001 when $(\lambda, \beta) = (.10, .10)$ and .007 when $(\lambda, \beta) = (.65, .30)$, corresponding to a range between 60 and 423 thousands infected people as of June 20, 2020.

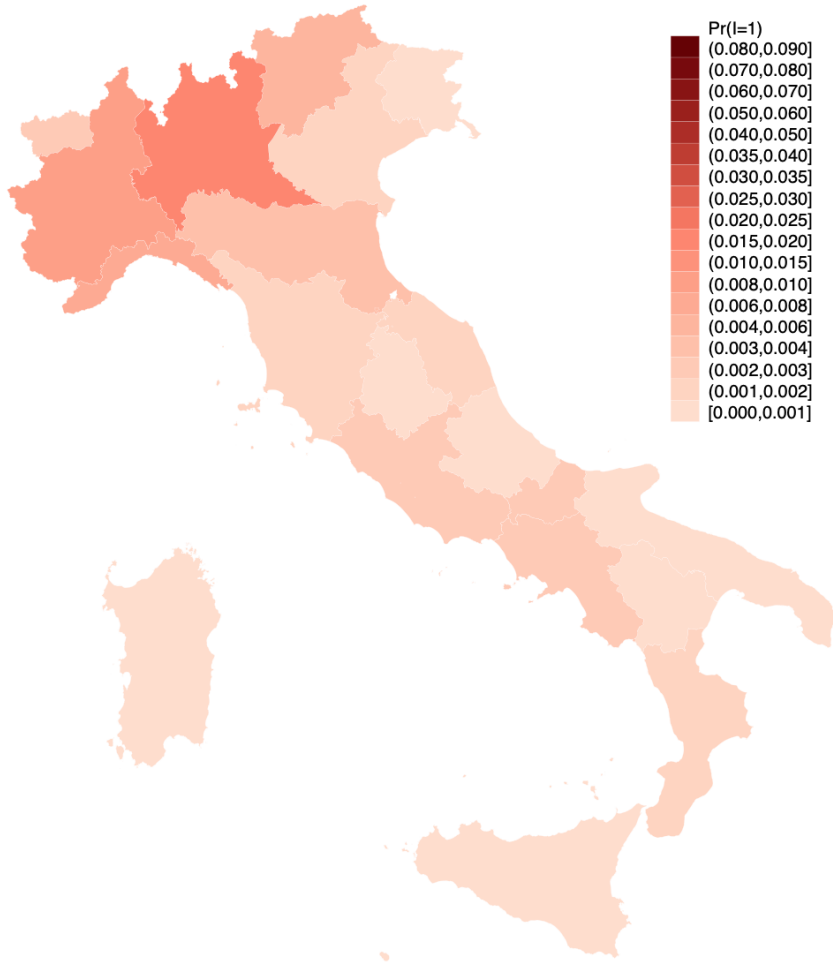
Remarks

- Table 2: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β), June 20, 2020.
- $\mathbb{P}(I = 1)$ increases with both λ and β, ranging between 0 for $(\lambda, \beta) = (0, 0)$ and .015 for $(\lambda, \beta) = (.99, .50)$.
- Given the population size of 60.4 million, the upper bound corresponds to about 900 thousands infected people as of June 20, 2020.
- In red, estimated $\mathbb{P}(I = 1)$ for the restricted range $.10 \leq \lambda \leq .65$ and $.10 \leq \beta \leq .30$.
- Estimated $\mathbb{P}(I = 1)$ ranges between .001 when $(\lambda, \beta) = (.10, .10)$ and .007 when $(\lambda, \beta) = (.65, .30)$, corresponding to a range between 60 and 423 thousands infected people as of June 20, 2020.

Table 3: Estimated $\mathbb{P}(I = 1)$ for different values of (λ, β)

Region	(λ, β) (.25, .10)	(.50, .20)	(.75, .30)
Abruzzo	.000	.000	.001
Basilicata	.000	.000	.000
Calabria	.001	.001	.002
Campania	.001	.003	.004
Emilia-Romagna	.002	.004	.007
Friuli Venezia Giulia	.000	.001	.001
Lazio	.001	.002	.004
Liguria	.003	.007	.011
Lombardia	.008	.017	.029
Marche	.001	.002	.003
Molise	.001	.002	.004
Piemonte	.004	.009	.016
Puglia	.000	.001	.001
Sardegna	.000	.000	.001
Sicilia	.000	.000	.001
Toscana	.001	.002	.003
Trentino-Alto Adige	.002	.004	.007
Umbria	.000	.000	.000
Valle d'Aosta	.001	.002	.004
Veneto	.001	.001	.002
Italy	.002	.005	.008

Figure 1: Estimated $\mathbb{P}(I = 1)$ by region



λ = .50, β = .20, Day = 6/20/2020

Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time

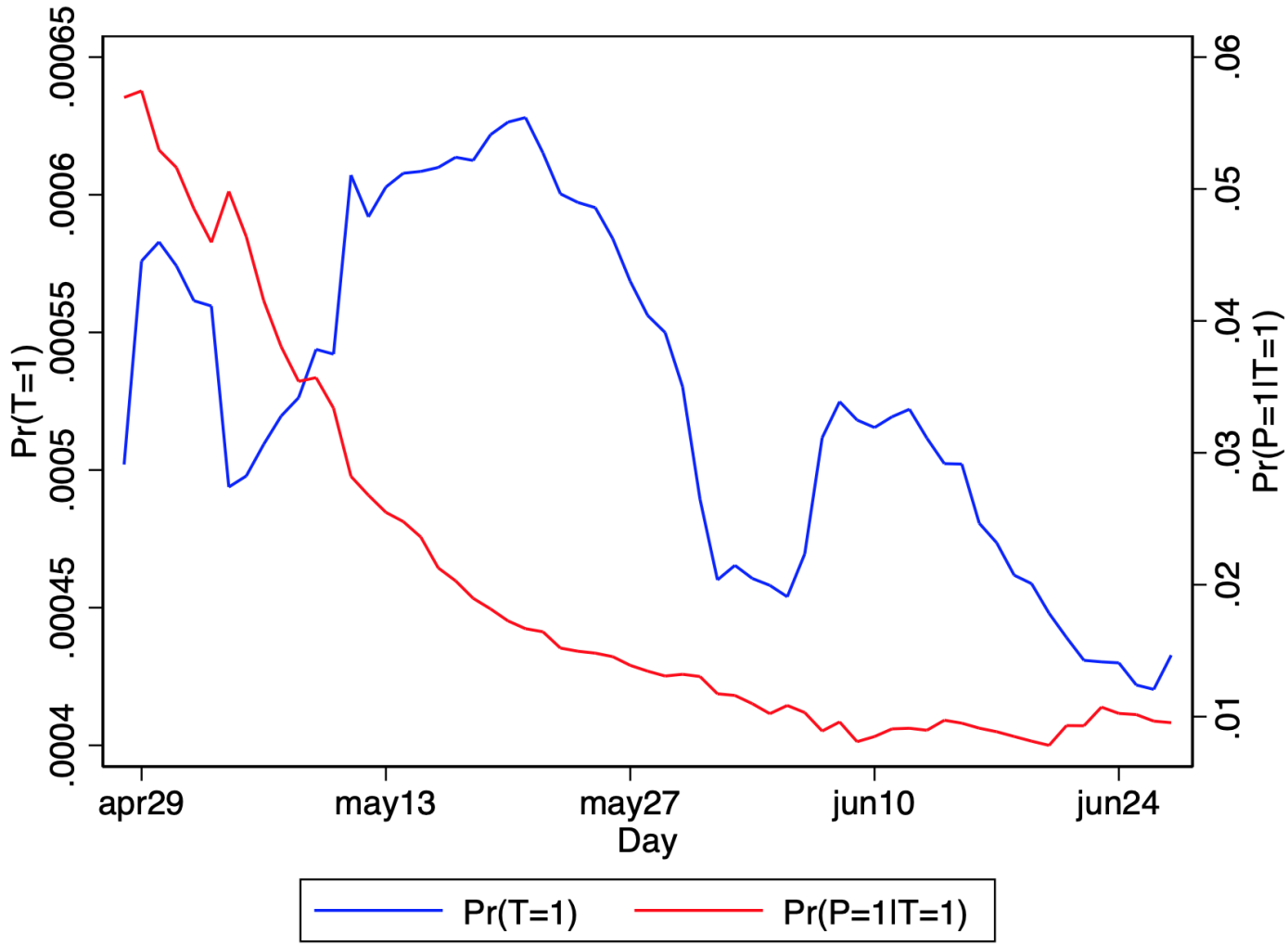


Figure 3: Estimated $\mathbb{P}(I = 1)$ over time

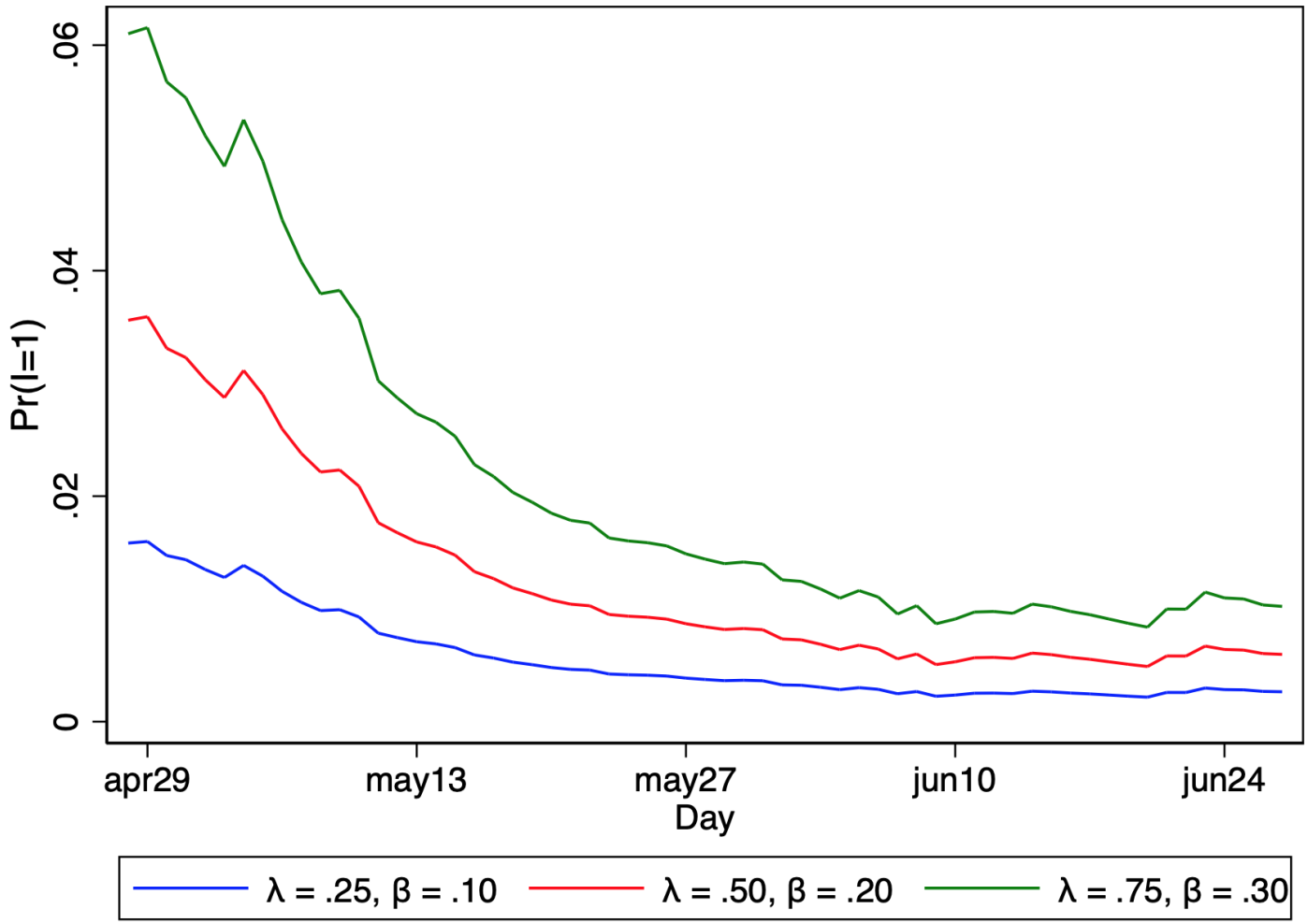
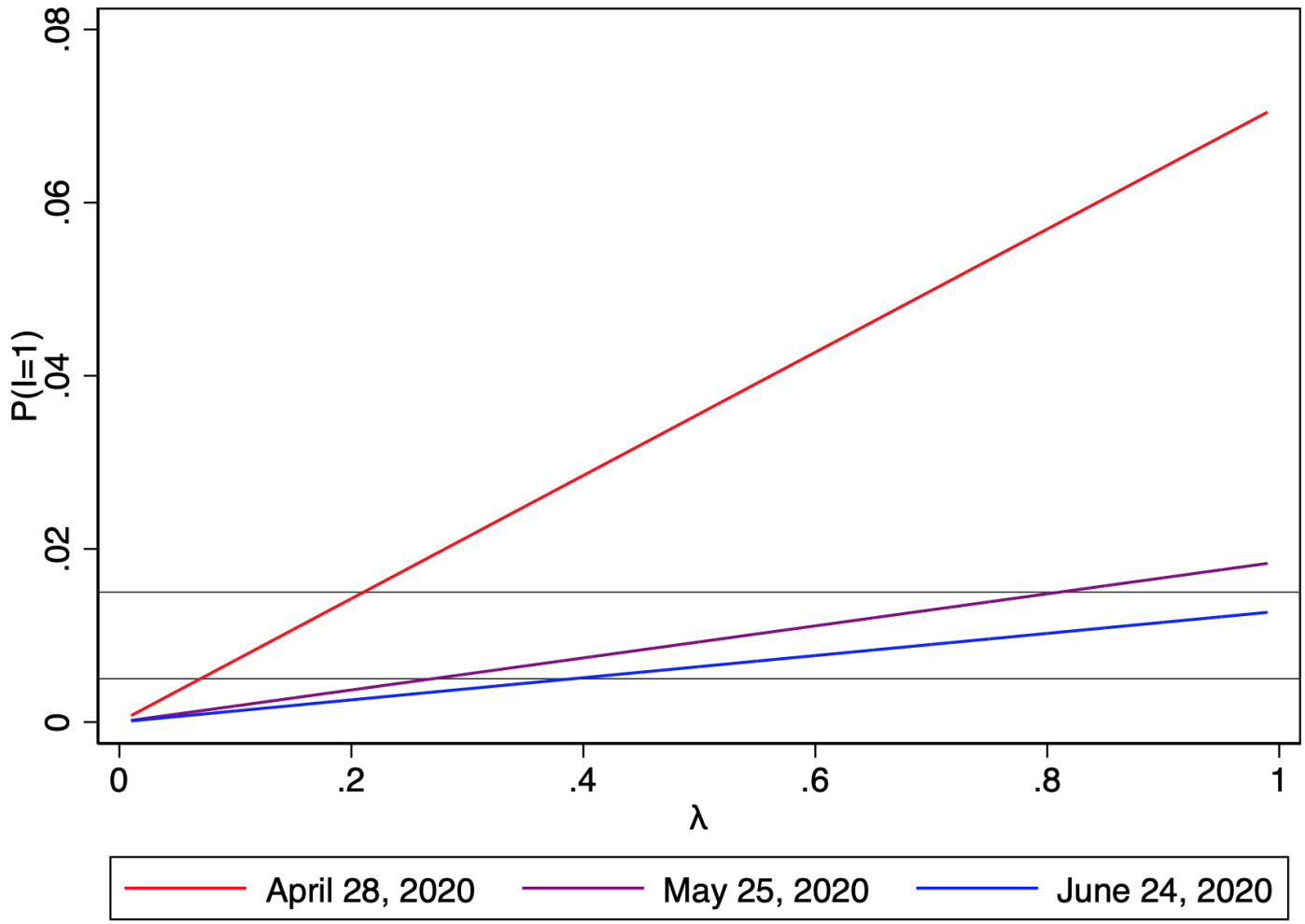


Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$



Remarks

- Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time.
- Figure 3: Estimated $\mathbb{P}(I = 1)$ over time.
- The observed decline reflects the fact that, while β is likely constant, $\mathbb{P}(T = 1)$ and especially $\mathbb{P}(P = 1|T = 1)$ fall.
- It may also reflect a shrinking testing bias (increasing values of γ), which implies λ increasing towards 1.
- Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$. The horizontal lines correspond to .015 and .005, the estimated values of $\mathbb{P}(I = 1)$ in mid-May and mid-June.

Remarks

- Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time.
- Figure 3: Estimated $\mathbb{P}(I = 1)$ over time.
- The observed decline reflects the fact that, while β is likely constant, $\mathbb{P}(T = 1)$ and especially $\mathbb{P}(P = 1|T = 1)$ fall.
- It may also reflect a shrinking testing bias (increasing values of γ), which implies λ increasing towards 1.
- Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$. The horizontal lines correspond to .015 and .005, the estimated values of $\mathbb{P}(I = 1)$ in mid-May and mid-June.

Remarks

- Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time.
- Figure 3: Estimated $\mathbb{P}(I = 1)$ over time.
- The observed decline reflects the fact that, while β is likely constant, $\mathbb{P}(T = 1)$ and especially $\mathbb{P}(P = 1|T = 1)$ fall.
- It may also reflect a shrinking testing bias (increasing values of γ), which implies λ increasing towards 1.
- Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$. The horizontal lines correspond to .015 and .005, the estimated values of $\mathbb{P}(I = 1)$ in mid-May and mid-June.

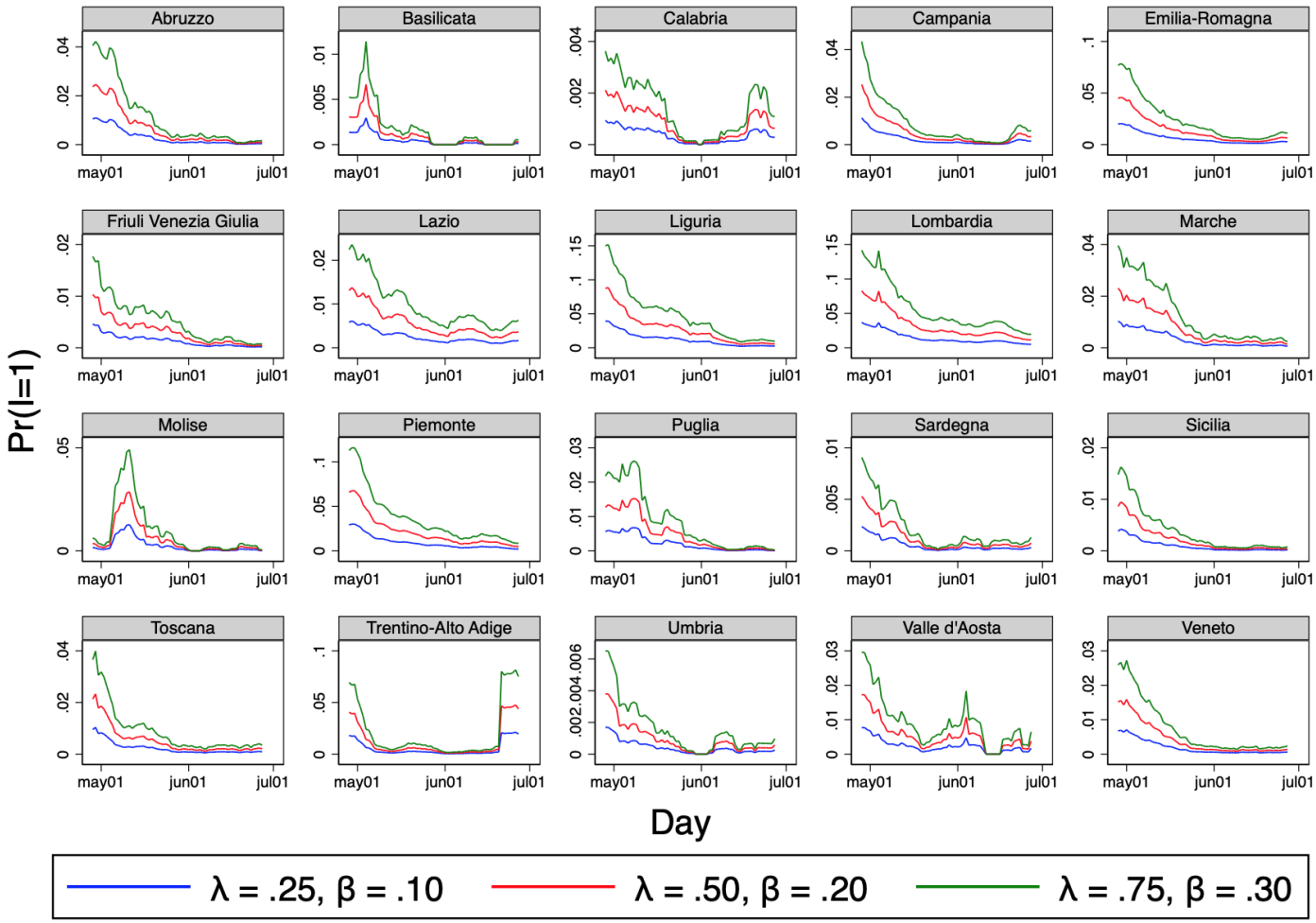
Remarks

- Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time.
- Figure 3: Estimated $\mathbb{P}(I = 1)$ over time.
- The observed decline reflects the fact that, while β is likely constant, $\mathbb{P}(T = 1)$ and especially $\mathbb{P}(P = 1|T = 1)$ fall.
- It may also reflect a shrinking testing bias (increasing values of γ), which implies λ increasing towards 1.
- Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$. The horizontal lines correspond to .015 and .005, the estimated values of $\mathbb{P}(I = 1)$ in mid-May and mid-June.

Remarks

- Figure 2: $\mathbb{P}(T = 1)$ and $\mathbb{P}(P = 1|T = 1)$ over time.
- Figure 3: Estimated $\mathbb{P}(I = 1)$ over time.
- The observed decline reflects the fact that, while β is likely constant, $\mathbb{P}(T = 1)$ and especially $\mathbb{P}(P = 1|T = 1)$ fall.
- It may also reflect a shrinking testing bias (increasing values of γ), which implies λ increasing towards 1.
- Figure 4: Estimated $\mathbb{P}(I = 1)$ at different dates for $\beta = .20$. The horizontal lines correspond to .015 and .005, the estimated values of $\mathbb{P}(I = 1)$ in mid-May and mid-June.

Figure 5: Estimated $\mathbb{P}(I = 1)$ by region and over time



Remarks

- Figure 5: Estimated $\mathbb{P}(I = 1)$ by region and over time.
- Erratic behavior of $\mathbb{P}(I = 1)$ in regions with a small number of cases (Basilicata, Calabria, Molise, Trentino-Alto Adige, Valle d'Aosta).

Remarks

- Figure 5: Estimated $\mathbb{P}(I = 1)$ by region and over time.
- Erratic behavior of $\mathbb{P}(I = 1)$ in regions with a small number of cases (Basilicata, Calabria, Molise, Trentino-Alto Adige, Valle d'Aosta).

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.

Conclusions

- We showed how to use readily available data to obtain quick estimates of the point prevalence of the infection, an important gauge of its evolution.
- Our estimates depend on a few clearly identified features of the epidemic, about which it is possible to gather information. The latter, in turn, can be used to narrow the range of the estimates.
- We also showed that another important statistic – the period prevalence of the infection – is harder to pin down with the data currently available, and we clarified which additional information is needed to obtain such an estimate.
- Finally, we applied our method to data from Italy.
- In some Italian regions the point prevalence might be still large enough to suggest caution in removing all restrictions on mobility.
- The health authorities, by incorporating their more detailed and disaggregated information into our framework, could obtain a quickly updated gauge of the evolution of the infection, which might help them to contain it.