OPTIMIZING A COMPARTMENT ODE MODEL AND NUMERICAL VERIFICATION

Jeremy Chiu,
APMA 923, Dr. JF Williams

December 2023

1 Project Overview

We consider the compartment ODE model presented by Guo-Wu[2], which describes the spread of Tuberculosis (TB) among the Canadian foreign-born population. Their ODE model features 5 dependent variables (each requiring an initial condition) and 9 parameters, some of which are physically difficult to measure or estimate. Our goal is to find the combination of parameters and initial conditions that give the global minimum of the model's prediction error, which is defined by comparing model output to reported TB incidence data.

1.1 Background and loss function

We consider the compartment ODE model presented by Guo-Wu[2], which describes the spread of Tuberculosis (TB) among the Canadian foreign-born population. Fundamentally, an SEIR-model is used, where the *exposed* category is partitioned into *early latent* and *late latent*. Figure 1 shows an overview of their model.

After fixing initial conditions and parameter values, we use Matlab's `ode23` routine to solve the system of differential equations from [2], which returns population vs time. From the population, we compute the TB incidence

$$\text{Estimated TB Incidence } = \frac{100,000}{X(t)+E(t)+L(t)+T(t)+R(t)} \cdot (pwE(t)+vL(t)),$$

which can be compared to the reported TB incidence from Canadian reports[3] (see figure 2). We use data from 2010-2020. The error is defined to be the difference between our computed (estimated) incidence and the reported incidence. We use parameters that are difficult to measure as input to the optimizer: q_1 and q_2 are the percentage of new immigrants that are undetected by Canadian TB screening, and the initial populations E_0, L_0, and R_0 are unknown; for a total of 5 variables

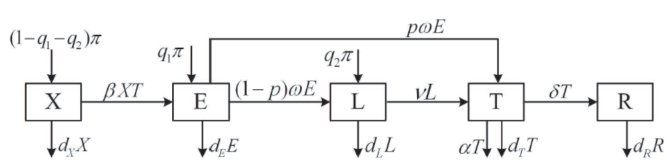

Figure 1: Flow diagram of the compartment ODE model used by [2].

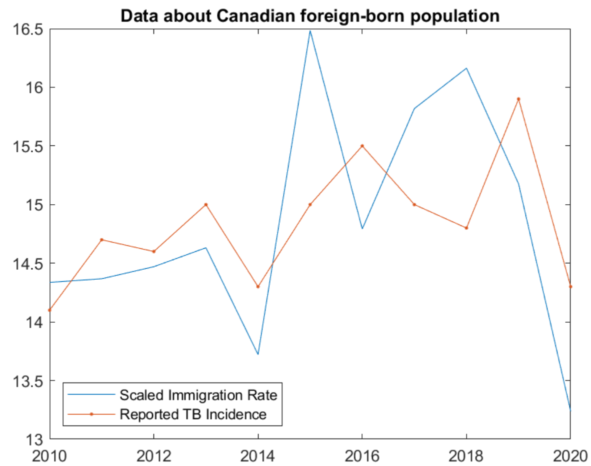

Figure 2: Reported incidence of Canadian foreign-born population. Annual immigration rate is also presented, scaled to be on same viewing rectangle. Notice how the number of immigrants drive the incidence rate.

to optimize across. Note that T_0 is reported, and X_0 is computed from the reported total foreign-born population. All other parameters use estimates found from the medical and epidemiological literature.

1.2 Project Proposal Meeting

JF and I met on October 27 to discuss the project. Here were JF's suggestions about possible things to try:

- Do not over expand on problem background. Discuss how we solved it.
- Compare solvers.
- Try to find local sensitivity. Investigate derivatives with respect to parameters.
- Discuss `fmincon`.
- Global optimization toolbox.
- Find an upper bound on q_1 and q_2 using immigration data.
- Investigate convexity.

2 Finding the global min with `fmincon`

We used Matlab's constrained solver, `fmincon`. Our constraints were

$$q_1,\ q_2,\ E_0,\ L_0,\ R_0 \geq 0,\ \text{and}\ q_1 + q_2 \leq 1$$

Initial conditions were the original proportions presented in [2]. The solver does a decent job in minimization – figure 3 shows the estimated incidence after optimization, which is reasonably close to the reported incidence.

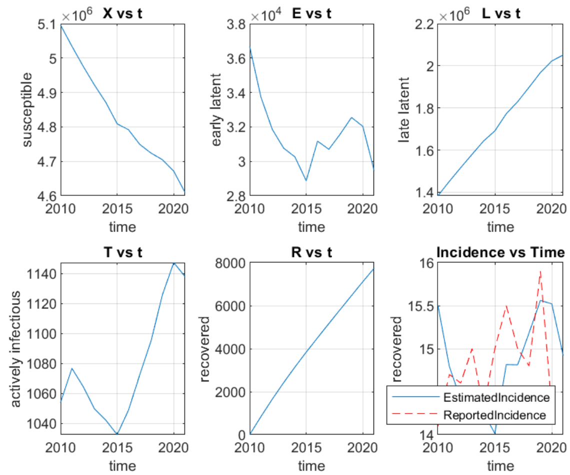

Figure 3: After optimization, the model reasonably estimate incidence.

2.1 Sensitivity analysis

Aside from the 5 variables we optimized across, we had various parameters that we fixed from the literature review. We explored the impact these parameters had on our model by simulating over a range of their values. In total, over 2000 combinations were tried, and so the simulations must be carefully stored and interpreted.

As demonstrated in figure 4, we found that the results are most sensitive to ν, ω, and L_0; this may be unsurprising since they explicitly appear in the loss function. However, although E_0 appears in the loss function, our model is not sensitive to its choice.

2.1.1 Histogram and two issues

We also explored the results of the sensitivity analysis by looking at the distribution of optimized variables. Figure 5 (a) gives a histogram for the optimized estimates of each of q_1, q_2, E_0, L_0 and R_0. The histogram demonstrates two issues:

1. Many optimal solutions are near $(q_1, q_2) = (0, 0)$.
2. The optimal L_0 and R_0 are often very similar to its initial choice.

Issue 1) Figure 5 (a) shows many optimal q_1 and q_2 are nearly 0. $q_2 = 0$ is non-realistic, because it would mean there are no immigrants with latent TB, which is unlikely because latent TB cannot be detected by many TB-infectivity tests. Moreover, some immigrants come from countries where more than half the population is infected with TB, and our screening process targets identifying people with actively infectious TB. For now, this issue can be avoided by discarding the non-realistic solutions. Going forward, we should be careful with how we select initial conditions that result in $(q_1, q_2) = (0, 0)$.

Issue 2) is more problematic. The optimizer is failing to look across L_0 and R_0. The optimal solution being near the initial condition means whatever we specify as the initial condition becomes the optimal solution. This would force us to have accurate measurements of L_0 and R_0 to gain realistic results; unfortunately, L_0 and R_0 are difficult to measure in the real-world. Figure 5 (b) highlights how L_0 and R_0 are very frequently near the initial condition. We should aim to either non-dimensionalize and remove them from the optimizer, or find a way to accurately specify them. Another way to interpret issue 2 is that our optimizer is currently sensitive to the initial condition for L_0 and R_0.

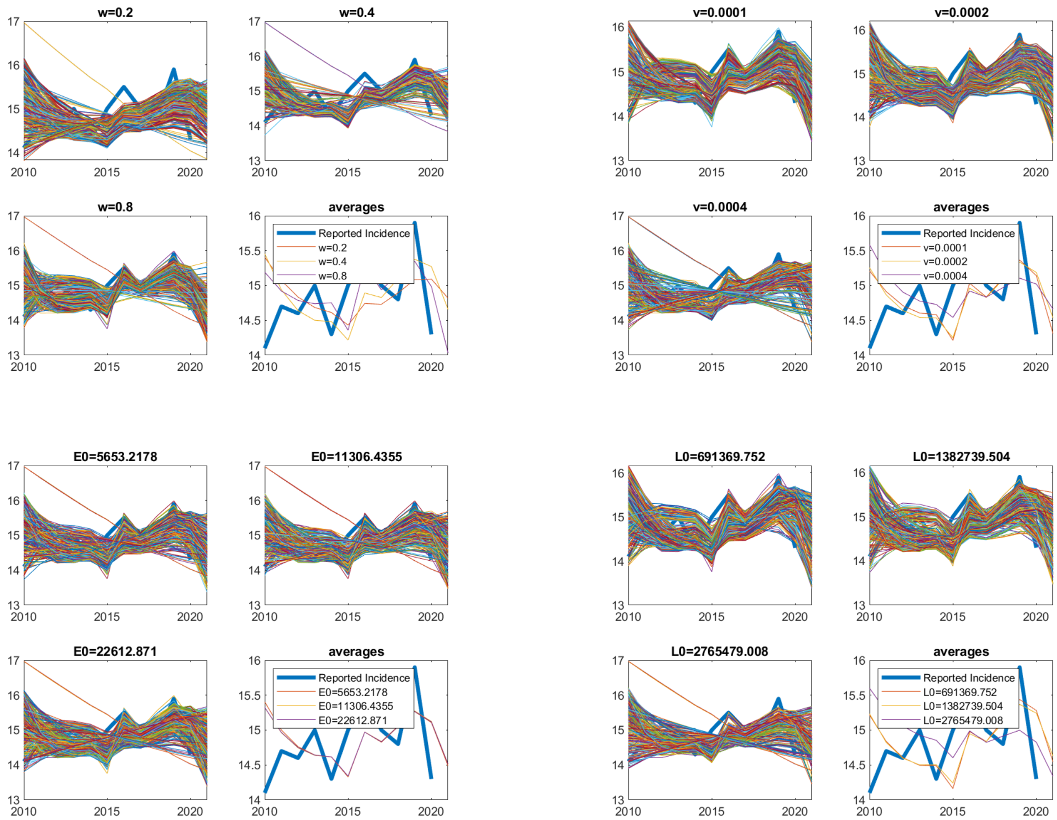

Figure 4: Our results were most sensitive across ν, ω, and $Ł_0$. There is little sensitivity across the choice of initial E_0.

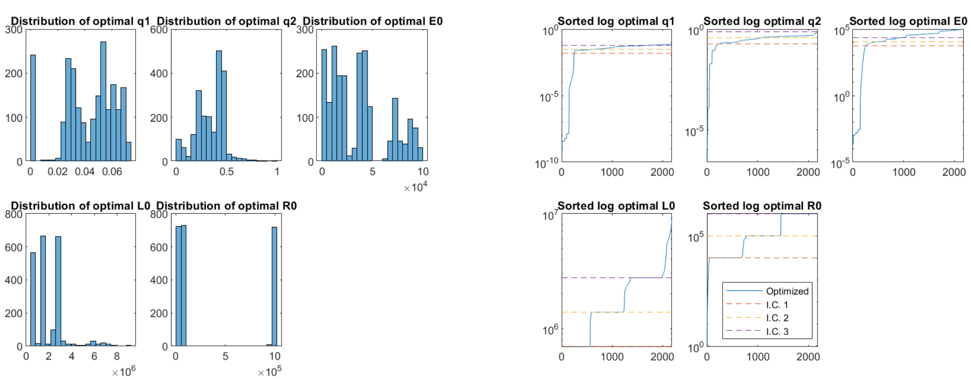

Figure 5: (a) Historgram of optimized parameters. Notice the L_0 and R_0 are similar to their initial value. (b) The optimal variables are sorted and plotted, along with the initial value. Notice how many of the optimal L_0 and R_0 are very similar to the initial condition. The optimal values were logged for a better viewing rectangle.

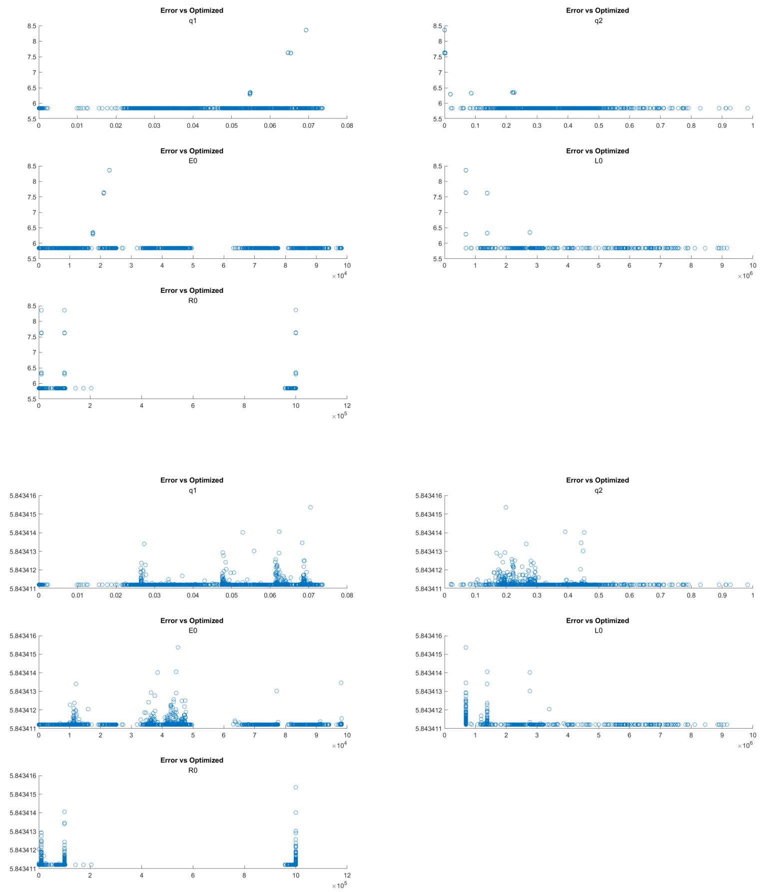

Figure 6: Scatter plot of the error vs optimized parameter. As we zoom in and delete outliers, we see more apparent outliers. Will Ruth: "This is what the tail of a distribution will often look like. It suggests that the outliers aren't typos".

2.1.2 Distribution of error vs optimal parameter

To get a sense of how the optimal solutions are performing, for each variable, we plot the error vs the output of the optimizer, presented in figure 6. We prune outliers and continue to zoom in, but see more and more apparent outliers. I spoke with Will Ruth (who just finished his PhD in statistics with Richard Lockhart) who commented, "This is what the tail of a distribution will often look like. It suggests that the outliers aren't typos."

In the scatter plot for R_0 in figure 6, notice the three vertical lines correspond to the initial conditions.

3 Steady state as initial condition

To address issue (2), which highlights the need for accurate initial specifications for L_0 and R_0, instead of using the initial conditions from [2], we simulate to steady states, and use the steady state as the initial condition.

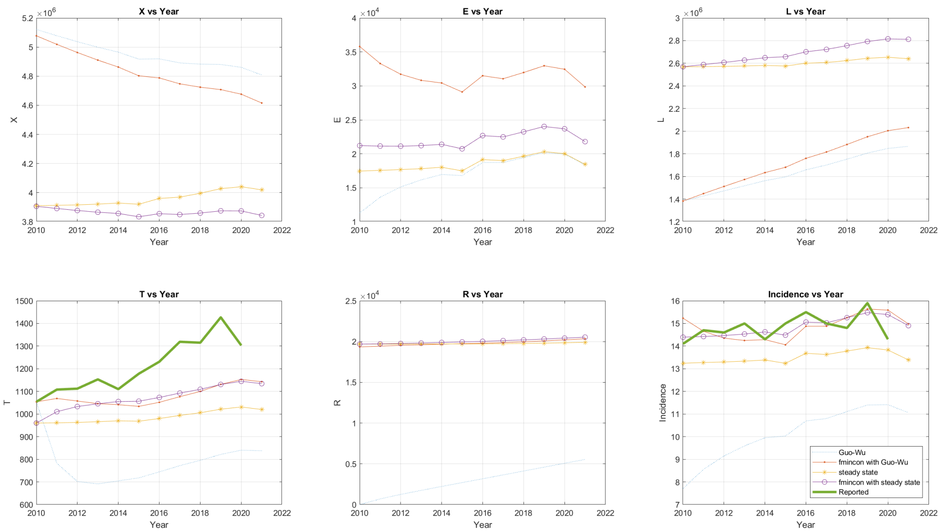

Figure 7: Even without optimization, the steady state is quite close to the reported incidence and prevalence.

3.1 Pointwise stationary approximation

In queuing theory, the pointwise stationary approximation of a queue system approximates any time of a queue system using its steady state (holding any time-dependent parameters constant) [1].

Borrowing this idea, we estimate the initial conditions using the steady state of the system of DEs, which has a globally asymptotically stable unique endemic equilibrium [2]. We fix the time-dependent parameter immigration rate π at reported value in 2010, then numerically simulate to steady state. We have three ways to assess the accuracy of this approximation:

1. Compare the steady state's total population to the reported total population.
2. Compare the steady state's T_0, which is reported to be 1054.
3. Compare the steady state's estimated incidence, which is reported to be 14.1.

By these metrics, the steady state does reasonably well; see figure 7. This suggests the steady state can be used, either to help specify the initial condition, or to remove variables from the set of inputs.

3.2 Sensitivity analysis on L_0 and R_0

To study the effects of using steady states as initial conditions, we perform sensitivity analysis across various values of E_0, L_0, and R_0. Sensitive parameters (ν and ω) are held constant so we can ignore their effects for now. Figure 8 shows our optimizer is no longer sensitive to the initial condition of L_0 and R_0, which is good.

Earlier, our solver had issues with the initial conditions of L_0 and R_0. Now, our results are most sensitive to q_1 and q_2, suggesting we are ready to pick realistic values of q_1 and q_2 to proceed.

4 Sensitivity analysis on ω, ν, q_1, and q_2

Uses steady state as initial condition.

From earlier investigations, we found our results were most sensitive to ω, ν, q_1, and q_2. We picked an initial condition for them, then scaled them by $(0.90)^{[1,0,-1,-2]}$, giving a total of $4^4 = 256$

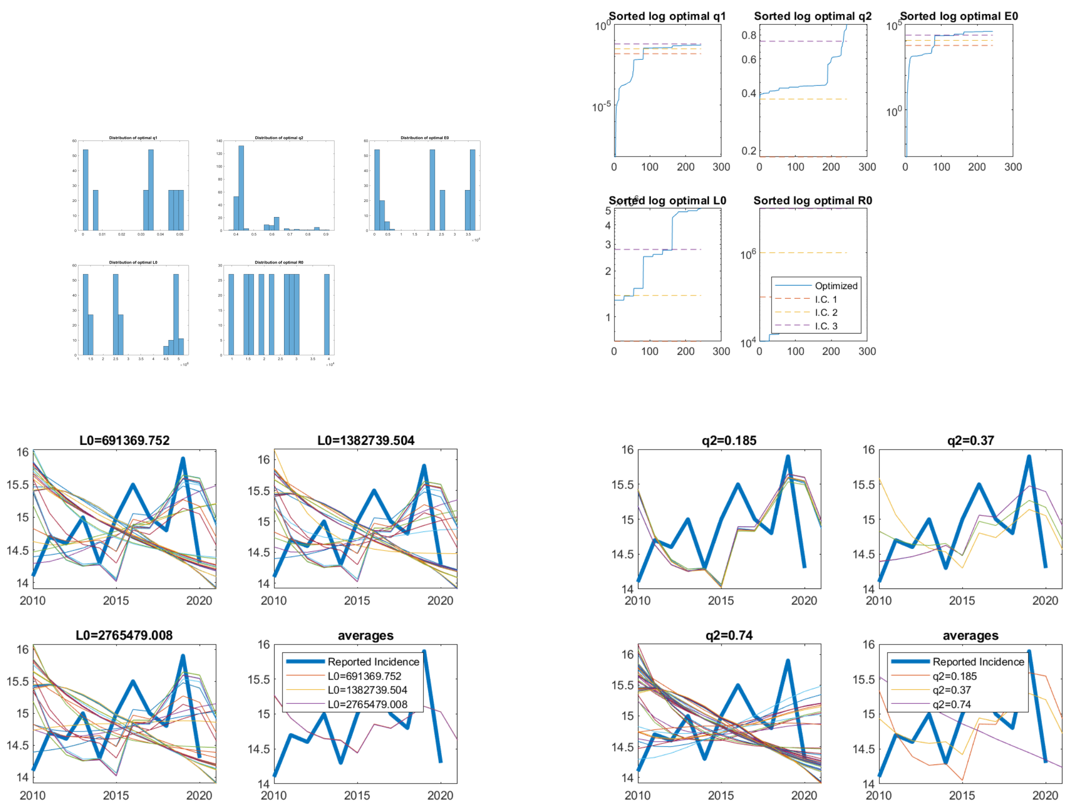

Figure 8: After updating initial conditions to be steady states, the optimal results were not dependent on L_0 or R_0. Instead, the response was most sensitive to q_1 and q_2. We held ν and ω constant.

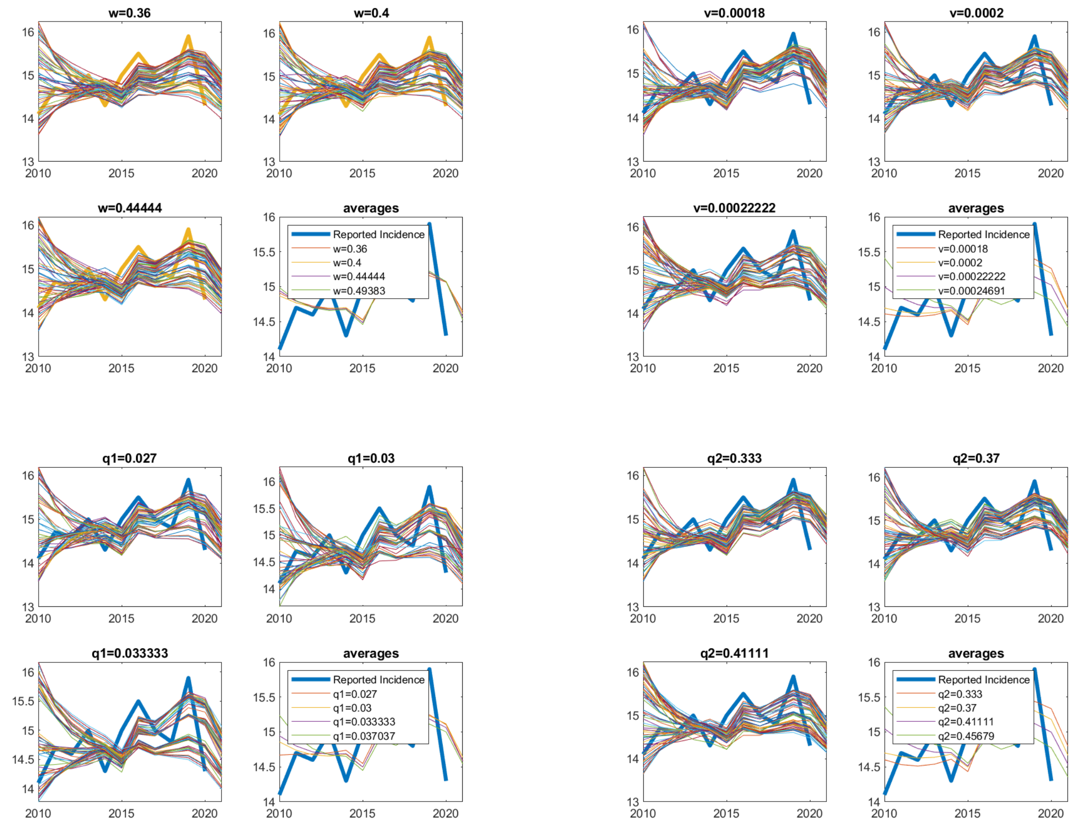

Figure 9: I don't see anything irregular with these results. The only things I notice is that as ν gets smaller, the incidence in 2010 improves.

combinations. See figure 9, which shows the results are not sensitive to ω, but somewhat sensitive to ν, q_1, and q_2. I suspect ν is the most important because q_1 and q_2 are thrown into the optimizer anyway.

Figure 10 shows the histogram of the optimal parameters. L_0 is trimodal. Otherwise, the other distributions look reasonably smooth.

4.1 Examining `fmincon` output

Among the 256 experiments we performed during sensitivity analysis, we look at the set of parameters that minimize and maximize the errors, then examine some of `fmincon`'s outputs. For the minimizing set, `fmincon` took 11 iterations with a total function count of 142. This number is consistent with the fact that `fmincon` uses central finite differencing; since there are 5 dimensions, each gradient would cost 10 function evaluations (i.e. 110 function evaluations to evaluate the gradient). Other factors like backtracking could contribute to the remaining 32 evaluations.

4.2 Gradient and Hessian - preconditioning. scaling variables

We found

$$\nabla f_{\min} = [4.46e-03,\ -1.56e-05,\ -3.13e-10,\ -7.99e-12,\ 2.76e-12],$$

and

$$\nabla f_{\max} = [4.09e-01,\ -5.00e-03,\ -1.21e-03,\ -9.43e-06,\ 1.35e-09].$$

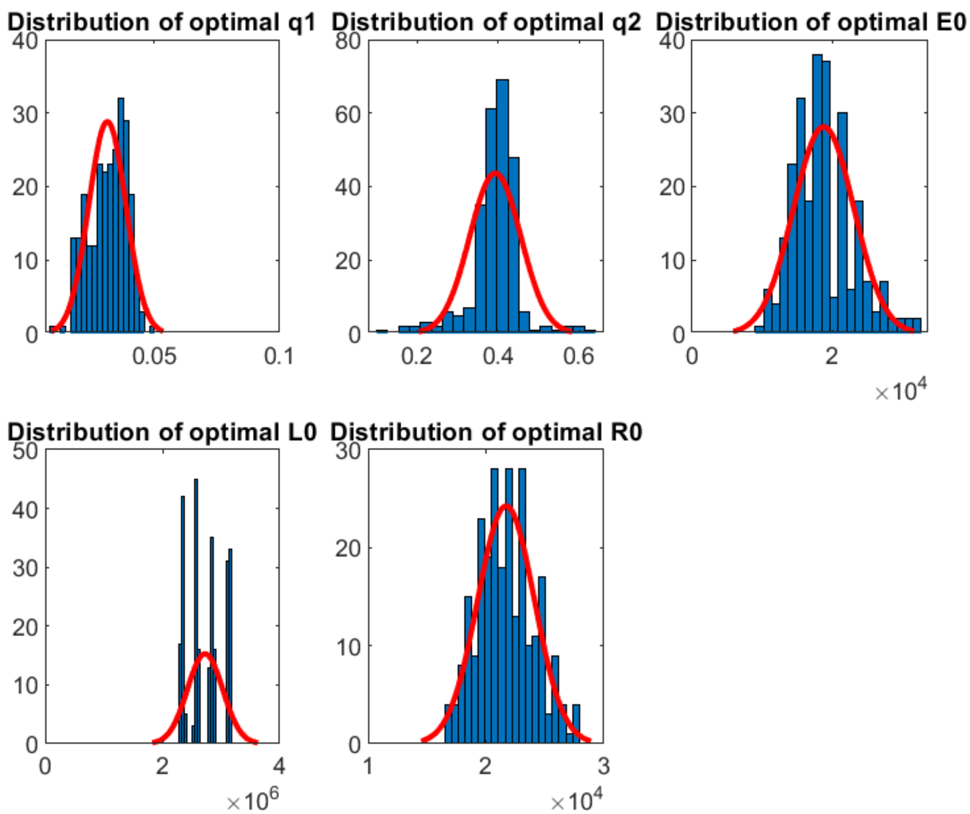

Figure 10: The most alarming histogram is the tri-modal L_0. Only one initial condition for L_0 was specified.

Notice the largest partial derivatives are f_{q_1} and f_{q_2}, while f_{E_0}, f_{L_0}, and f_{R_0} are small. This is consistent with our sensitivity analysis demonstrating that our model was most sensitive to q_1 and q_2.

Recall that we expect the gradient to approach zero near a minimum. Notice that

$$\|\nabla f_{\min}\|_\infty = 4.46e - 03 \ll \|\nabla f_{\max}\|_\infty = 4.09e - 01.$$

$\nabla f_{\max}$ is 100 times larger than $\nabla f_{\min}$, suggesting the the optimizer should continue searching at the maximum. It turns out the error-maximizing parameter is poorly conditioned, which is demonstrated by the Hessian H.

4.3 Hessian

Recall at a local minimum, the Hessian should be symmetric positive definite. This is numerically verified by computing the eigenvalues of the Hessian matrix. We found the eigenvalues λ

$$\lambda(H_{\min}) = [5.2868e - 04,\ 3.2415e - 01,\ 1.0000e + 00,\ 1.0000e + 00,\ 8.4916e + 05],$$

while

$$\lambda(H_{\max}) = [2.9168e - 09,\ 1.0000e + 00,\ 1.0000e + 00,\ 1.5888e + 02,\ 1.7682e + 08].$$

Although all eigenvalues are positive, the small eigenvalues in $H_{\max}$ are driving the condition number up, which may be causing problems for the optimizer and thus the large gradient.

For $H_{\min}$, the eigenvectors are as follows (each column is an eigenvector):

$$\begin{bmatrix} -0.0314146 & -0.0003089 & 0.0000034 & 0.0000000 & -0.9995064 \\ 0.9994670 & 0.0088770 & -0.0001204 & -0.0000016 & -0.0314161 \\ 0.0067183 & -0.7473443 & 0.6637665 & -0.0290771 & 0.0000221 \\ -0.0000319 & 0.0044353 & 0.0487478 & 0.9988013 & -0.0000002 \\ 0.0058115 & -0.6643628 & -0.7463496 & 0.0393770 & 0.0000201 \end{bmatrix}$$

with corresponding eigenvectors

$$\begin{bmatrix} 0.0005287 & 0.3241506 & 1.0000000 & 1.0000000 & 849155.2834507 \end{bmatrix}$$

The largest eigenvalue's eigenvector is dominated by the term corresponding to q_1. This suggests each variable is on a different order of magnitude, and so we should rescale.

4.3.1 Taylor Expansion

Recall the multivariate Taylor expansion of a function $f(x)$ at a point $a \in \mathbb{R}^n$:

$$y = f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^{\mathrm{T}}(\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^{\mathrm{T}}\mathbf{H}(\mathbf{a})(\mathbf{x} - \mathbf{a}) + O(\|\mathbf{x} - \mathbf{a}\|^3).$$

At a minimum, the gradient is zero, and so the function is approximately quadratic and described by the Hessian. I'm not sure how eigenvectors and eigenvalues play into this.

5 Normalization of x

For each $y \in \vec{x}$ (so y is either $q_1,\ q_2,\ E_0,\ L_0,\ R_0$), we scale by $\frac{y - y_{\min}}{y_{\max} - y_{\min}}$. The min and max are the the smallest and largest optimized values from sensitivity analysis.

Results in figure 11

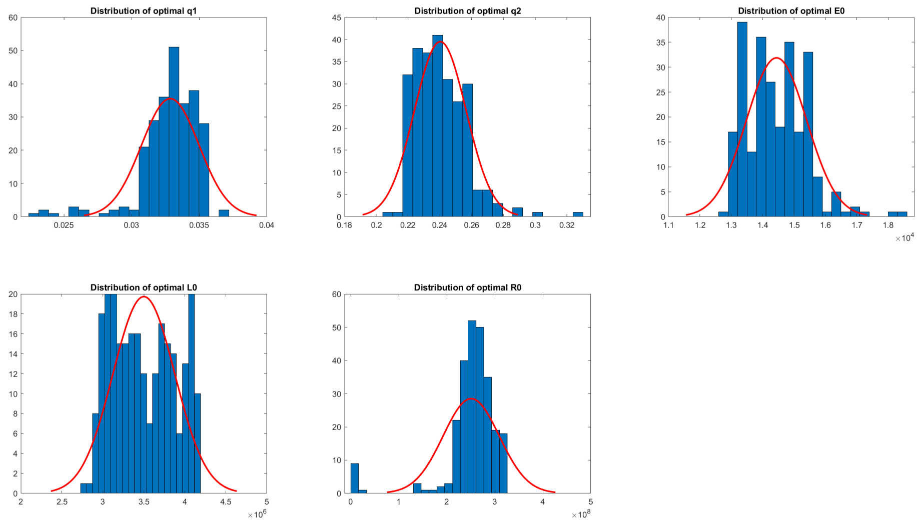

Figure 11: Normalized by scaling (y-ymin)/(ymax-ymin).

6 Future work and other attempts

1. *Global optimization toolbox.* I coded `multistart`, but found the solutions were still sensitive to the initial conditions we specified.

2. From figure 2, we see that immigration rate drives incidence. Currently our initial condition is set in 2010, but perhaps it should be in 2009 instead (i.e., a year before the data set). The immigration rate $\pi(2009)$ would likely be helpful.

3. Investigate the multi-modal nature of L_0.

4. Newly infected patients (≤ 2 years from infection) are $15\times$ more likely to develop active TB than people with no known risk factor [4], hence

$$p\omega = 15 \cdot \nu.$$

   This can be compared to our choices of ω and ν.

5. Will's suggestion: Aim to reduce variability among optimal q_1 and E_0. Among all the optimal q_1, we can measure variability by computing standard deviation divided by its mean.

6. Estimate derivative analytically. Computing partial derivatives gets complicated because the loss function involves an integral of a high-dimensional function. The main benefit, according to Sandy and JF, was savings in computational time (`fmincon` uses finite differencing to estimate derivative), but otherwise would not help our optimizer much.

7. Find upper bound by looking at actual immigration data. We've currently collected a breakdown of which countries Canadian immigrants arrived from for the years 2016-2020 and are continuing to expand this collection. Afterward, we could bucket countries into high-incidence and low-incidence, which can help improve estimates our of q_1 and q_2.

7 Meeting Notes and Questions

Jan. 2

- Discuss Taylor expansion. At minimum, gradient is zero, so approximately quadratic. Look at Hessian
- Eigenvalues and eigenvectors of Hessian. The eigenvectors are nearly basis vectors. Some eigenvalues big, some large. We should scale to fix this.
- For each $y \in \vec{x}$ (so y is either $q_1,\ q_2,\ E_0,\ L_0,\ R_0$),

  we scale by $\frac{y-y_{\min}}{y_{\max}-y_{\min}}$. The min and max are the the smallest and largest optimized values from sensitivity analysis.

Jan.9

1. Albert mentioned we should look at solution. How do I plot it?
2. I see Hessian in Taylor expansion. I don't see how eigenvectors matter.

post-meeting:

- if we want a global incidence to Canada from the world (q_1 and q_2), fraction of immigrants dot product with incidence from within each country
- 0-1 normalization ; normalization of variables
- bi-modalness. around R0, there are zero; hand full of runs trying to estimate 0. check if these outliers correspond to outliers in q1 q2 E0 L0
- decision variable
- fix all decision variables except say q1. plot objective function vs q1.
- profile of an objective function is choose one of variables, optimize all the others.
- do these numbers make sense. check bounds.
- if i start at a, and i walk along one of the eigenvectors, the eigenvalues tell me how much I change along the direction of that eigenvector. direction of maximal change is the eigenvector with largest eigenvalue; lucky interpretable
- Fischer information
- maximum likelihood analysis may help us. calculate likelhood, maximize it, that's our estimator. there is limit theory and observation theroy, as observations go to infinity we can describe distribution of the maximum likelhood estimator

  for starters, consistency - the max likelhood est convergence [of random objects, be careful; in probability, in other contexts] to true parameter value, modulus regularity conditions

  asymptotic normality. estimator converges to truth (consistency). so subtract the limit, we converge to 0. what if multiply by something going to infinity (at right rate), can get non-trivial finite limit. right rate is sqrt of sample size; which converges to normal distribution. mean is zero (not suprrising, subtract by limit); variance of this limit depends (inverse) directly is the Hessian of the log likelihood. ergo, places wehre likelhood has lots of curvature are places where Hessian is large, which means iunverse hessian is small, meaning in limit MLE (max likelihood estimator) has small variance. ergo large curvature implies small variance.
- variables are on different ranges, which would induce difference in curvature

decision variable

References

[1] Linda V. Green and Peter J. Kolesar. The pointwise stationary approximation for queues with nonstationary arrivals. *Management Science*, 37:84–97, 1991.

[2] Hongbin Guo and Jianhong Wu. Persistent high incidence of tuberculosis among immigrants in a low-incidence country: Impact of immigrants with early or late latency. *Mathematical Biosciences and Engineering*, 8(3):695–709, 2011.

[3] Mounchili A., Perera R., Henry C., Soualhine H., Sharma M., Yoshida C., Rabb M., and M. Carew. *Tuberculosis Surveillance in Canada 2010-2020 Summary Report.* Public Health Agency of Canada, 2022.

[4] Public Health Agency of Canada. *Canadian Tuberculosis Standards.*, volume 7th Edition. Centre for Communicable Diseases and Infection Control (Canada), 2014.