43. **Tichonov estimate with model uncertainty.** (p.91). We wish to choose the slope, s, of a linear scalar model:

$$y = sx \tag{130}$$

We have a prior estimate of the slope, $\widetilde{s}$, and we have data, (x_i, y_i), $i = 1, \ldots, M$. The Tichonov estimate of s minimizes:

$$T = \lambda(\widetilde{s} - s)^2 + (1-\lambda)\frac{1}{M}\sum_{i=1}^{M}(y_i - sx_i)^2 \tag{131}$$

where $0 \le \lambda \le 1$. We will assume that x and y are dimensionless quantities.[4]

(a) Derive an expression for the estimate of s which minimizes T.

(b) Now consider model uncertainty, with two different info-gap models:

$$\mathcal{U}(h) = \{y = sx + u : \ |u| \le h\}, \quad h \ge 0 \tag{132}$$

$$\mathcal{U}(h) = \left\{y = sx + ux^2 : \ |u| \le h\right\}, \quad h \ge 0 \tag{133}$$

For each info-gap model, derive an expression for the robustness of an estimate of the slope. How does the robust-satisficing estimate differ between the two models? How do they differ from the Tichonov estimate? Note that, because x and y are dimensionless, the horizons of uncertainty in these two info-gap models are also dimensionless. This makes the robustnesses which are evaluated with these two info-gap models comparable.[5]

[4] If x and y have units, and even if they have the same units, then the units of T in eq.(131) are undefined. This means that the relative weights of the two terms in T are controlled by the units, not by the value of λ.

[5] If x and y have units then it is be necessary to calibrate the two robustnesses, which requires judgment and cannot be done uniquely. However, if x and y have units then we also face a different problem, noted in footnote 4.

Solution for problem 43. (p.25)

(a) We find the Tichonov optimal slope by differentiation:

$$\frac{\partial T(s)}{\partial s} = -2\lambda(\widetilde{s}-s) - \frac{2(1-\lambda)}{M}\sum_{i=1}^{M}(y_i - sx_i)x_i \tag{756}$$

$$= \left(2\lambda + \frac{2(1-\lambda)}{M}x^T x\right)s - 2\lambda\widetilde{s} - \frac{2(1-\lambda)}{M}x^T y \tag{757}$$

The 2nd derivative is positive so the solution is a minimum. Thus the Tichonov optimal slope is obtained by equating eq.(757) to zero:

$$s_{\mathrm{T}} = \frac{\lambda\widetilde{s} + [(1-\lambda)/M]x^T y}{\lambda + [(1-\lambda)/M]x^T x} \tag{758}$$

We see that, if $\lambda = 1$ then $s_{\mathrm{T}} = \widetilde{s}$. Likewise, if $\lambda = 0$ then s_{T} equals the ordinary least-squares optimal slope.

(b) Let $T(s,u)$ denote the Tichonov error with model uncertainty. The robustness of slope s, with requirement of error no greater than T_{c}, is defined as:

$$\widehat{h}(s, T_{\mathrm{c}}) = \max\left\{h : \left(\max_{u\in\mathcal{U}(h)} T(s,u)\right) \le T_{\mathrm{c}}\right\} \tag{759}$$

Let $\mu(h)$ denote the inner maximum. We now consider the two different info-gap models.

Info-gap model of eq.(132), p.25. The Tichonov error is:

$$T(s,u) = \lambda(\widetilde{s}-s)^2 + \frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i - u)^2 \tag{760}$$

$$= \underbrace{\lambda(\widetilde{s}-s)^2 + \frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i)^2}_{T_o(s)} - 2u\frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i) + (1-\lambda)u^2 \tag{761}$$

where $T_o(s)$ is the ordinary Tichonov error.

We find $\mu(h)$ at $u = -h\,\mathrm{sgn}\left(\sum_{i=1}^{M}(y_i - sx_i)\right)$. Thus:

$$\mu(h) = T_o(s) + 2h\frac{1-\lambda}{M}\left|\sum_{i=1}^{M}(y_i - sx_i)\right| + (1-\lambda)h^2 \tag{762}$$

$$= T_o(s) + 2h(1-\lambda)\left|\overline{y} - s\overline{x}\right| + (1-\lambda)h^2 \tag{763}$$

This is the inverse of the robustness curve: a plot of h vs. $\mu(h)$ is the same as a plot of $\widehat{h}(s, T_{\mathrm{c}})$ vs. T_{c}.

Consider two different slope values, one of them the Tichonov optimum s_{T}. Let $\mu(h,s)$ and $\mu(h, s_{\mathrm{T}})$ denote the inverse robustnesses, eq.(763). Equating $\mu(h,s) = \mu(h,s_{\mathrm{T}})$ we see that their robustness curves cross at the following value of robustness if it is positive:

$$h_{\times} = \frac{T_o(s) - T_o(s_{\mathrm{T}})}{2(1-\lambda)(|\overline{y} - s_{\mathrm{T}}\overline{x}| - |\overline{y} - s\overline{x}|)} \tag{764}$$

By definition, $T_o(s_{\mathrm{T}}) \le T_o(s)$, so the numerator is non-negative. If $\lambda < 1$ and $s_{\mathrm{T}} \ne \overline{y}/\overline{x}$, then one can choose s in infinitely ways so that the denominator is also positive. In other words, given these two conditions, we have proven that there are an infinite number of non-Tichonov slopes which are more robust than s_{T} for all values of T_{c} exceeding a threshold.

Info-gap model of eq.(133), p.25. The Tichonov error is:

$$
\begin{aligned}
T(s,u) &= \lambda(\widetilde{s}-s)^2 + \frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i - ux_i^2)^2 && (765)\\
&= \underbrace{\lambda(\widetilde{s}-s)^2 + \frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i)^2}_{T_o(s)} - 2u\frac{1-\lambda}{M}\sum_{i=1}^{M}(y_i - sx_i)x_i^2 + u^2\frac{1-\lambda}{M}\sum_{i=1}^{M}x_i^4 && (766)
\end{aligned}
$$

where $T_o(s)$ is the ordinary Tichonov error as before.

We find $\mu(h)$ at $u = -h\,\mathrm{sgn}\left(\sum_{i=1}^{M}(y_i - sx_i)x_i^2\right)$. Thus:

$$
\begin{aligned}
\mu(h) &= T_o(s) + 2h\frac{1-\lambda}{M}\left|\sum_{i=1}^{M}(y_i - sx_i)x_i^2\right| + h^2\frac{1-\lambda}{M}\sum_{i=1}^{M}x_i^4 && (767)\\
&= T_o(s) + 2h(1-\lambda)\left|\overline{yx^2} - s\overline{x^3}\right| + h^2(1-\lambda)\overline{x^4} && (768)
\end{aligned}
$$

This is the inverse of the robustness curve: a plot of h vs. $\mu(h)$ is the same as a plot of $\widehat{h}(s, T_{\rm c})$ vs. $T_{\rm c}$.

We can again argue, like in eq.(764), that there are an infinity of non-Tichonov slopes which are more robust than $s_{\rm T}$. Evaluating eq.(768) for s and for $s_{\rm T}$ and solving for h we find that the robustness curves cross at:

$$
h_\times = \frac{T_o(s) - T_o(s_{\rm T})}{2(1-\lambda)\left(\left|\overline{yx^2} - s_{\rm T}\overline{x^3}\right| - \left|\overline{yx^2} - s\overline{x^3}\right|\right)} \qquad (769)
$$

This is positive if $\lambda < 1$ and $s_{\rm T} \neq \overline{yx^2}/\overline{x^3}$.

Comparing the robustnesses of the two info-gap models. The robustnesses are directly commensurable because they are both dimensionless. We ask: is one model more robust than the other, or do their robustness curves cross one another?

Let us denote the inverse robustnesses for these two info-gap models, eqs.(763) and (768), by $\mu_1(h)$ and $\mu_2(h)$ respectively.

Note that:

$$
\mu_1(0) = T_o(s) = \mu_2(0) \qquad (770)
$$

This means that the robustness curves sprout off of the $T_{\rm c}$ axis at the same value, $T_o(s)$.

Do the robustness curves cross again at positive robustness, or is one robust-dominant?

Equate $\mu_1(h) = \mu_2(h)$ and solve for h. The robustness curves cross at positive robustness if and only if there is a positive solution:

$$
h_\times = \frac{2\left(|\overline{y} - s\overline{x}| - \left|\overline{yx^2} - s\overline{x^3}\right|\right)}{\overline{x^4} - 1} \qquad (771)
$$

This may positive.