Motivating Example

- Suppose we are trying to fit a production frontier, but have little information about its functional form:

$$Y_i = F(X_i)\Delta_i\tilde{\varepsilon}_i$$

  - Y_i is ith observation of output
  - X_i is ith observation of inputs
  - F is (unknown) production function
  - $\Delta_i \in [0,1]$ is technical efficiency
  - $\tilde{\varepsilon}_i > 0$ is observation error

Motivating Example: Data

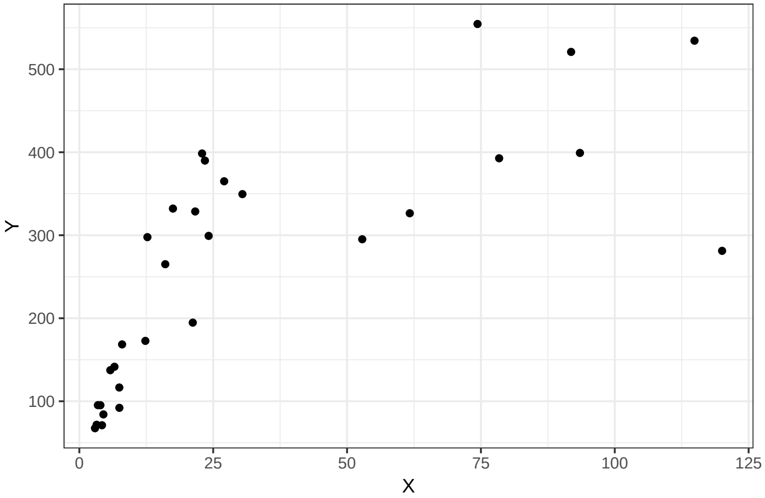

Figure: Simulated Data

Motivating Example: Cleaning Data

- Common first step: Log-transform:

$$y_i = f(X_i) + \delta_i + \varepsilon_i$$

  - $y_i = \log Y_i$
  - $f(X_i) = \log F(X_i)$
  - $\delta_i = \log \Delta_i \leq 0$
  - $\varepsilon_i = \log \tilde{\varepsilon}_i$
- Make distributional assumptions:
  - $\delta_i \sim N^{-}(0, \sigma_\delta)$
  - $\varepsilon_i \sim N(0, \sigma_\varepsilon)$

Motivating Example: Data

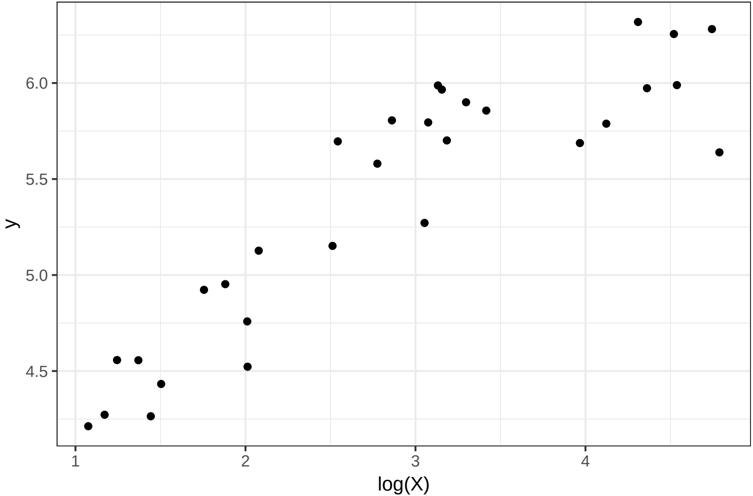

Figure: Log Simulated Data

Motivating Example: Functional Forms

- We still need to select a form for f
- Traditionally, parametric forms of f have been used
  - Log-linear: $f(X_i) = \log(X_i)\beta$
  - Translog: All log-inputs, squared log-inputs, and interactions between log-inputs
- Could also use a non-parametric specification
  - Du et al. (2013) use kernel smoothing to fit conditional mean of the data, use residuals to fit distributional parameters
- How can we select between these functional forms, especially given limited data?
  - Log-linear and translog forms can be compared using many methods
  - Difficult to compare with non-parametric methods

Motivating Example: Functional Forms

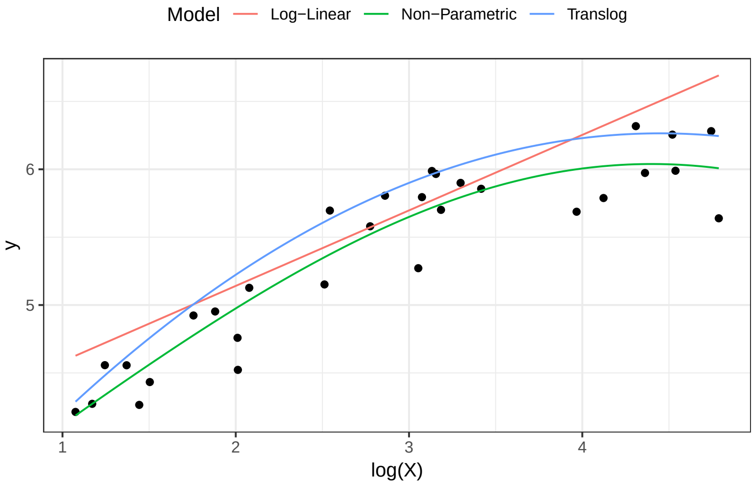

Figure: Fitted Frontiers

Model Selection

- Classical methods
  - Based on likelihood values
  - Some methods are restrictive in what models can be compared (e.g., nested models), others like information criteria are general
  - Require moderate to large sample sizes
  - **Problems:**
    - Kernel smoothing is not likelihood-based
    - Sample sizes may not be large enough
    - Classical methods are unreliable for estimating stochastic frontiers

Model Selection

- Cross-validation
  - Fit the model with one sub-sample, test accuracy against another sub-sample
  - Requires large sample sizes and a metric for accuracy
  - **Problem:** Sample sizes are not large enough

Model Selection

- Bayesian methods
  - Very general model comparison, can estimate the probability a model is correct from a set of exclusive models
    - Accounts for likelihood and numbers of parameters
  - No sample size restrictions in general
  - **Difficulties:**
    - Need a Bayesian analogue of kernel smoothing (Gaussian processes)
    - Calculating model probabilities is hard in general, existing methods are unsuitable for large models applied to small samples

Bayesian Model Selection

- Given a set of exclusive models $\{M_1, ..., M_K\}$, the probability that model M_k is the true model given data y is

$$\Pr(M_k|y) = \frac{m(y|M_k)p(M_k)}{\sum_{j=1}^{K} m(y|M_j)p(M_j)}$$

  - $m(y|M_k)$ is marginal likelihood of model M_k
  - $p(M_k)$ is prior probability that M_k is the true model
- Marginal likelihoods are difficult to compute in general

Marginal Likelihood Estimation

- Many methods have been developed to compute marginal likelihoods
  - If Gibbs or Metropolis-Hastings sampling are used, efficient methods developed in Chib (1995) and Chib and Jeliazkov (2001)
  - Laplace's method and Gaussian quadrature perform well for moderately-sized models
  - Bridge sampling is very efficient and is widely applicable (in theory)

Marginal Likelihood Estimation: Problems

- Existing methods are not well suited to marginal likelihood estimation for this model and data
  - Model is not of correct form for Gibbs sampling
  - Excessive convergence times for Metropolis-Hastings sampling
  - Models have too many parameters relative to sample size for Laplace's method, accuracy is degraded
  - Models have too many parameters for Gaussian quadrature to be computationally feasible
  - Bridge sampling is prone to numerical issues in large models, accuracy is degraded (more on this later)

Numerical Issues With Bridge Sampling

- Bridge sampling estimates marginal likelihood with

$$\hat{m}(y) = \frac{\frac{1}{N_2}\sum_{s=1}^{N_2} p\left(y|\theta_g^{[s]}\right) p\left(\theta_g^{[s]}\right) h\left(\theta_g^{[s]}\right)}{\frac{1}{N_1}\sum_{s=1}^{N_1} h\left(\theta_y^{[s]}\right) g\left(\theta_y^{[s]}\right)}$$

  - θ denote parameters
  - $p(y|\theta)$ is the likelihood
  - $p(\theta)$ are priors over parameters
  - $g(\theta)$ is the proposal distribution, chosen by the researcher
  - $h(\theta) = (r_1 p(y|\theta)p(\theta) + r_2 \hat{m}(y)g(\theta))^{-1}$
  - θ_g are parameter samples taken from g (total of N_2 samples)
  - θ_y are parameter samples taken from the posterior distribution (total of N_1 samples)

Numerical Issues With Bridge Sampling

$$\hat{m}(y) = \frac{\frac{1}{N_2}\sum_{s=1}^{N_2} p\left(y|\theta_g^{[s]}\right) p\left(\theta_g^{[s]}\right) h\left(\theta_g^{[s]}\right)}{\frac{1}{N_1}\sum_{s=1}^{N_1} h\left(\theta_y^{[s]}\right) g\left(\theta_y^{[s]}\right)}$$

- For large numbers of parameters, $p(\theta)$ and $g(\theta)$ can be very small positive numbers
- Thus, $h(\theta)$ can take on very large values
- Terms in each sum can be very large or very small, are eventually truncated because of finite machine precision, making numerator and denominator inaccurate
- Inaccuracies are magnified by division of numerator by denominator
- Will show an example of biased marginal likelihood estimates that result