Estimation of Yield Densities: A Bayesian Nonparametric Perspective

***Yang Wang*:**

North Calorina State University, Department of Economics, 2801 Founders Drive, 4102 Nelson Hall, Raleigh, NC 27695. Email: ywang75@ncsu.edu.

***Francis Annan*:**

Columbia University, School of International and Public Affairs, 420 West 118th Street, New York, NY 10027. Email: fa2316@columbia.edu.

Selected Paper prepared for presentation at the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28

Estimation of Yield Densities: A Bayesian Nonparametric Perspective

Yang Wang[1], Francis Annan[2]

[1] Department of Economics, North Carolina State University [2] School of International and Public Affairs, Columbia University

Motivation

The statistical modeling of densities to capture yield risks is of great importance for rating various crop insurance products. In general, the small probability lower tail events within the distribution of yield outcomes trigger insurance payments. Such risky probabilities are however hard to quantify in practice due to the infrequent nature of these lower tail events. To this end, we propose and examine a Bayesian nonparametric model which is based on Dirichlet processes (DP) mixtures that provides potential gains in estimating yield densities.

Objectives

1. Propose a Bayesian nonparametric model which is based on DP mixtures for yield estimation.
2. Deploy our DP model for the empirical estimation of county level yield data for Cotton from Texas.
3. Examine the implications of our modeling framework on the pricing of the Group Risk Plan (GRP) insurance compared to a nonparametric Kernel-type model.

Data

The proposed model is deployed to estimate yield densities using annual Cotton yield data for 8 selected counties from Southern High Plains district in Texas. The yields are derived from publicly available county level data on total production and acres planted that come from National Agricultural Statistics Services (NASS) at the USDA. The yield series are constructed as production per acres planted to capture potential events that would trigger insurance payments. The data spans 1968-2013.

Methods

Follow 3 Steps

Step 1: Model the effects of technology on yields using seminonparametric approximation.

Step 2: Formulate a probability model for yield densities using mixtures of Dirichlet processes (DP).

Step 3: Derive inference using Markov Chain Monte Carlo (MCMC).

Model

First: we model the distribution of adjusted yields Y_t county-by-county in a fully Bayesian nonparametric setting,

$$Y_t \sim f(.), t = 1, \ldots, 46$$

Where $f(.)$ is the density of the yield. Formally we model the density using mixtures of DP which is specified as,

$$f(.|G) = \int \emptyset(Y_t|\mu,\sigma^2)\, dG(\mu,\sigma^2)$$

$$G|\alpha, G_0 \sim DP(\alpha G_0)$$

Where $\emptyset(.|\mu,\sigma^2)$ denotes the density of the Gaussian distribution with mean μ and variance σ , and G is a probability measure defined on $R \times R^+$. The unknown distribution G receives a DP prior which is specified with a concentration parameter, α and a precision parameter G_0, for G. The almost surely discrete nature of the DP allows for ties among the $\theta = (\mu, \sigma^2)$ to permit expressing the DP prior in a stick-breaking form of Sethuraman (1994),

$$G(.) = \sum_{k=1}^{\infty} w_k\, \delta_{\theta_k}$$

Where $\delta(.)$ denotes the Dirac measure θ_k=(μ_k, δ_k^2) and $k \in N$. The weights $w_k \in (0,1)$ arises from a stick-breaking format. We can cast the probability model as an infinite mixture of the form,

$$f(.|G) = \int \emptyset(Y_t|\mu,\sigma^2)\, dG(\mu,\sigma^2)$$

$$= \sum_{k=1}^{\infty} w_k\, \emptyset(Y_t|\mu_k, \delta_k^2\,)$$

This formulation shows that the yield density is defined.

Next: one can define the insurance liability as, $Liab = \pi\tau p$ where the guarantee is defined as the product of expected yield π and the coverage level τ, and p corresponds to the price of the commodity. The actuarially fair premium is,

$$\widehat{Prem} = p \int_0^{T=\pi\tau} (T-Y)\,\hat{f}(y)dy$$

where $\hat{f}(y)$ corresponds the estimated yield distribution. The associated premium rate is $\widehat{Prem}/Liab$.

Results

Figure 1: Estimated Densities for Hale County

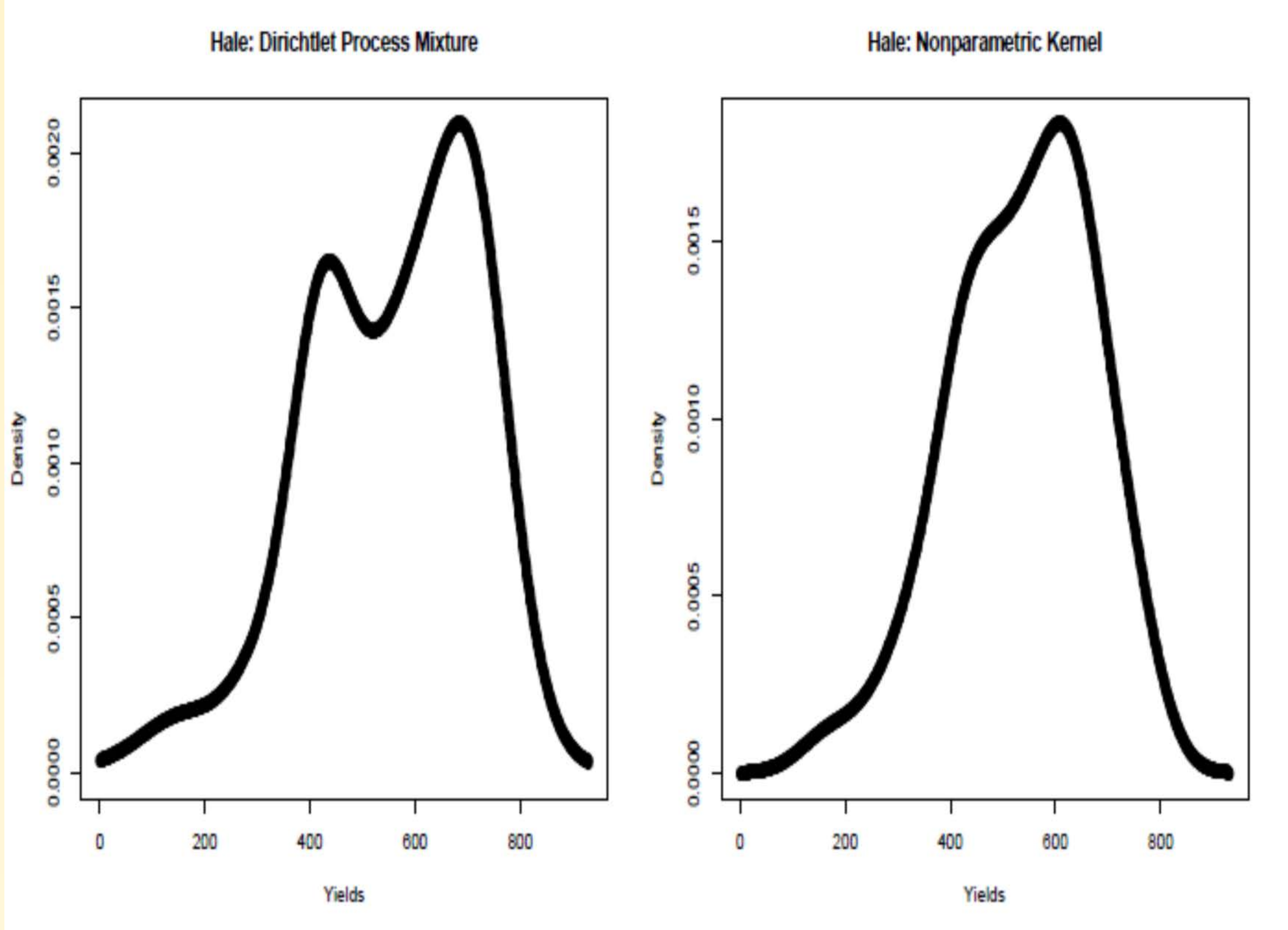
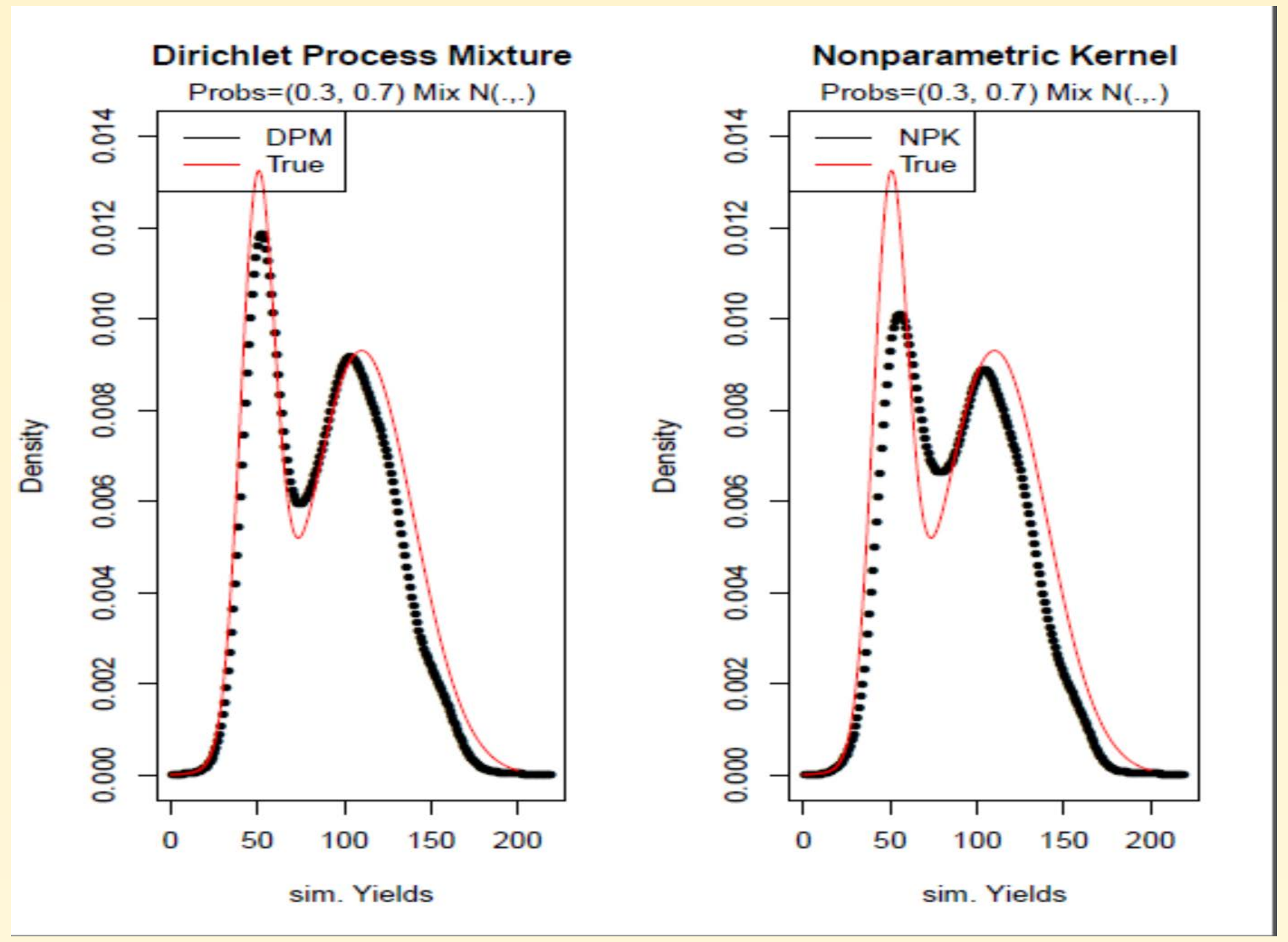


Table 1: **Actuarial estimates under alternative models**

County, TX	*Model*	*Exp. Loss Probs.*	*Premium ($)*	*Rate (%)*	*Equality Test*
Crosby					
	Dir.	.00046	17.08	4.28	
	Ker.	.00041	15.83	3.97	25.00 (0.00)
Dawson					
	Dir.	.00157	16.84	8.18	
	Ker.	.00103	14.84	7.21	27.32 (0.00)
Floyd					
	Dir.	.00041	30.08	5.98	
	Ker.	.00032	26.65	5.26	41.98 (0.00)
Gaines					
	Dir.	.00084	8.54	3.37	
	Ker.	.00068	8.46	3.33	16.47 (0.00)
Hale					
	Dir.	.00044	23.21	5.22	
	Ker.	.00036	21.68	4.87	29.54 (0.00)
Hockley					
	Dir.	.00070	15.88	4.85	
	Ker.	.00056	15.74	4.80	17.40 (0.00)
Lynn					
	Dir.	.00106	15.36	6.43	
	Ker.	.00081	13.61	5.70	28.95 (0.00)
Martin					
	Dir.	.00229	21.71	10.51	
	Ker.	.00145	17.24	10.34	27.86 (0.00)

Results

Figure 2: Model Fit Checks using Simulated Known Densities



1. The results provide visual evidence that the DP mixture estimate comparable distributions to the Kernel.
2. The left tail probabilities from DP mixture model are marginally larger, which reveal a relatively larger downside risks in the DP mixture model.
3. The DP mixtures model provides slightly higher insurance premium rates, which can translate into larger actuarial losses.
4. The DP mixtures model appears to fit known mixture of Gaussian densities better in the closeness check.
5. Results are robust to the selection of different Kernel bandwidth selection criteria.

Conclusions

The choice of density estimation approach could have important implications for the pricing of insurance. The Bayesian nonparametric approach provides a promising alternative of rating insurance products. This can be especially important in regions where the underlying yield distribution are substantially complex. The findings from this article will largely stimulate further research into application of the Bayesian nonparametric method.