Stochastic Frontier Yield Function Analysis to Predict Returns to a New Crop: An Example of Camelina *Sativa* Yields Conditional on Local Factor Levels

Sofia Kotsiri, Postdoctoral Research Scholar
Agricultural and Resource Economics
North Carolina State University
Email: skotsir@ncsu.edu

Kelly Zering, Associate Professor
Agricultural and Resource Economics
North Carolina State University
Email: kzering@ncsu.edu

Michelle Mayer, Research Associate
Agricultural and Resource Economics
North Carolina State University
Email: mlmayer@ncsu.edu

Selected Paper prepared for presentation at the Agricultural and Applied Economics Association's 2014 AAEA Annual Meeting, Minneapolis, MN, July 27-29, 2014

Abstract

The purpose of this study is to develop a model that calculates the probability distribution of camelina expected yields dependent on location-related variables such as precipitation, temperature, and solar radiation, as well as nitrogen rate and others. Camelina is an oilseed crop grown in cool climate with low input requirements including little water. The application to camelina addresses challenges in analysis of potential adoption of crops with limited field data. Our data include trials and crop yields in the United States from 2005 to 2012. They have been assembled from various published reports covering a range of locations, seasons, and production methods. We begin by fitting a least squares (LS) regression model to camelina yields. As a robustness check we also apply a stochastic frontier framework under Cobb-Douglas technology. Preliminary results indicate that the average maximum precipitation for the period of interest positively affected the mean camelina yields, whereas it has no impact on yield variability. An increase in average maximum precipitation will more likely decrease the technical inefficiency. Both higher nitrogen rates and higher average maximum growing degree days will more likely increase the average yields. A taller camelina plant positively affects the mean yields and the yield variability. In contrast, total solar radiation is negatively correlated with mean yields and variation. There is still much to be learned about the crop and its best management practices as production expands. The analysis of the interaction of managed input variables and environmental factors will help us assess varietal performance and provide location conditional predictions.

Introduction

Camelina *sativa* (also known as false flax or gold of pleasure) is a low input oilseed crop (Robinson, 1987), which appears to be promising in the race to find new biofuel crops. It is a short season cool climate crop with 85-100 days period of maturity, and can be grown on marginal land (implying minimal effect the food system) with low environmental footprint. So far, it has been produced at a larger scale in the Northwestern US (e.g., Colorado, Montana, Oregon, and North Dakota) and Saskatchewan and Alberta, Canada. Camelina is relatively robust such that yields are less dependent on weather conditions near harvest (Crowley and Fröhlich, 1998). Some authors suggest that it does not encounter pest problems faced by other oilseed crops, i.e., pollen beetle. It can be integrated in the fallow period of the crop rotation cycle (Johnson et al., 2009) thus reducing competition with other crops. Furthermore production costs are projected to be relatively low due to use of the same equipment as wheat (Keske et al., 2013, Newlands et al., 2012, Fullen and Polonius, 2011), and some suggest that low amounts of insecticide are required (Robinson, 1987). Gesch and Archer (2013) explored the profitability of double cropping soybeans with a winter type camelina in the upper Midwest. They found that combined oil-yields of double crop camelina and soybean exceeded the yields produced by mono-crops of soybean. Similarly, trials with camelina growing together with food crops such as grain and cereals proved that mixed cropping could bring up to 0.5 times higher relative yields in regions of Germany (Paulsed, 2011). However, some farmers prefer to remain with the traditional winter wheat and summer fallow because of limited availability of crop insurance for camelina[1], lack of USDA direct payments for camelina, as well as lack of knowledge and experience of

---

[1] Risk Management Agency (RMA) offers camelina crop insurance in Montana and North Dakota for $16/cwt but its duration is unknown.

growing this crop (Young et al., 2012). Furthermore, US camelina yields have been highly variable across time, location, and variety.

Camelina is a good source of oil as a fuel feedstock especially for the aviation sector, containing 30 percent to 40 percent oil. 14 major airlines airlines such as KLM, Lufthansa, and Japan Airlines as well as the US Navy have successfully operated flights with camelina derived jet fuel. The Department of Navy is a major driver of the derived demand for camelina since it plans to use eight billion gallons of renewable fuels by 2020 (Tindal, 2010). To achieve this goal, it would take over 70 million acres of crop production (AgriLogic LLG, 2011). By comparison, 88 million acres of corn were planted in the US in 2010 (USDA, NASS, 2010). Camelina oil can be either converted to biodiesel and used in a traditional diesel engine or burned directly in a modified diesel engine (Jewett, 2013). Furthermore after oil extraction, the remaining camelina seed can be sold as a high protein animal meal offering more revenue to camelina farmers (Brandess, 2012).

Our study tries to integrate biological models of individual plant growth with economic approaches. An NC State team is developing genetically modified camelina with improved characteristics and potentially higher yields. Agronomic development is already leading to improved camelina yield along with resistance to disease, weeds and pests (Moser, 2010). For example, in 2009, Agragen LLC introduced novel modifications that increase the tolerance of camelina to Group 2 herbicides by more than 300-fold in laboratory testing. Through an agreement with the University of Alberta and Agriculture and Agri-Food Canada (AAFC), Agragen will use its patented technology to introduce a gene encoding a key enzyme in oil synthesis for camelina (Glass, D. 2010). Zhang et al. (2012) found that under controlled environmental, overexpression of purple acid phosphatase 2 (AtPAP2) in camelina resulted in longer hypocotyls, earlier flowering, faster growth rate, higher photosynthetic rate and increased seed yield in comparison with the wild-type line.

Simulated crop rotation budget models (Brandess, 2012, Keske et al., 2013) have incorporated variables such as the cost of seeding, cost of nitrogen fertilizer, herbicide costs, the optimal camelina acreage, average yields, effect of rainfall on yield distribution (accounting for precipitation, as well as evapotranspiration and runoff), on-farm fuel needs, additional diesel fuel purchases, additional labor cost, interest on a six month operating loan, hauling, crushing, cleaning and filtering cost, storage, etc. Various biological crop simulation models have been developed (WEPP, EPIC, DSSAT, CropSyst, APSIM, etc.). We perform an econometric analysis of camelina yields using historical yield data and corresponding explanatory variables.

Conceptual Framework

We assume a production function $f(z_i, \beta)$. In the absence of inefficiency and error terms, the *i*th producer would produce

(1) $$q_i = f(z_i, \beta)$$

The stochastic frontier analysis assumes that each producer produces less than they might due to a degree of inefficiency (Kumbhakar and Lovell, 2000; Aigner, Lovell and Schmidt, 1977), i.e.:

(2) $$q_i = f(z_i, \beta)\tau_i,$$

where τ_i is the level of efficiency for the *i*th producer. If $\tau_i=1$, the producer is achieving the optimal output with the technology in $q_i = f(z_i, \beta)$, whereas for $\tau_i<1$ the producer is not making the most out of the inputs z_i. The τ_i is strictly positive and must be in the interval (0, 1].

Apart from the inefficiency, the output is subject to random shocks, thus it can be written as:

(3) $$q_i = f(z_i, \beta)\tau_i \exp(\upsilon_i)$$

The natural log of both sides yields:

(4)
$$\ln(q_i) = \ln\{f(z_i, \beta)\} + \ln(\tau_i) + \upsilon_i$$

Assuming *k* inputs and that the production function is linear in logs, defining $u_i = -\ln(\tau_i)$:

(5)
$$\ln(q_i) = \beta_o + \sum_{j=1}^{k} \beta_j \ln(z_{ji}) + \upsilon_i - u_i$$

Note that we only focus on technical inefficiency, and not allocative inefficiency that refers to the ability to combine inputs and outputs in optimal proportions after introducing prices of these inputs and output. (Lovell, 1993).

Empirical Model

Frontier models can be estimated either with stochastic frontier analysis that includes econometric methods or with Data Envelopment Analysis (DEA) that involves mathematical programming. Contrary to the stochastic frontier models, the DEA modeling does not require the specification of a functional form, and we cannot infer economic implications. Furthermore DEA is highly sensitive to outliers and the number of observations (Bravo-Ureta et al., 2007). We select the stochastic frontier since we seek to project camelina yields as conditional on climate variables that may significantly affect both the expected value and variability of output levels. Moreover, due to high collinearity, some variables may be excluded from the estimated form of the production function. For example, we exclude some variables used in previous analyses such as field size, labor use, and seeding rate.

We specify a Cobb-Douglas production function of the following form:

(6)
$$\ln Y_i = \beta_o + \beta_1 \ln(N) + \beta_2 \ln(GDD) + \beta_3 \ln(elev) + \beta_4 \ln(precip) + \beta_5 solar + \beta_6 height + \upsilon_i - u_i$$

where Y=camelina yields in pounds per acre, N=nitrogen rate applied, *GDD*=the average growing degree days for the period between planting and harvesting, *elev*=elevation , *precip*=average precipitation in inches for the period between planting and harvesting, *solar* =total solar radiation, *height*=camelina height in inches.

Data

Data on camelina production are limited since USDA started collecting crop data no earlier than 2007. Thus our data came mostly from published reports, University theses and Extension articles. We focus on the varieties that have been commercially released in the United States. The most common varieties are Blaine Creek and Suneson (Montana State University), Cheyenne (Blue Sun Biodiesel in Colorado) and Celina (developed in France). Jewett (2013) conducted a two year variety trial in the western United States and found that Ligena, SSD10, SSD177, SSD87, SSD138, and Celine were the highest yielding varieties in all combinations of environments, including irrigated environments. In our data set, the largest number of yield observations includes Calena, Ligena and then Blaine Creek, Suneson and Celina. Other common varieties are the BSX G22, BSX G24 (also from Blue Sun Biodiesel in Colorado) and Yellowstone (Great Plains Oil and Exploration in Ohio). Most of the spring seeding occurs between early April and early May and harvesting rangesfrom mid August to early September. Seed rates have not been found to significantly affect the yields, whereas nitrogen rate affected both yields and crop height (Crowley and Fröhlich, 1998). The total plot area in our data ranged between 75 sqft to 2,000 sqft.

Data for soil temperatures (not included in the analysis reported here) were collected from multiple sources since there is no unique database that contains soil temperature data for the sites and the dates we are interested in. Nitrogen rates were recorded as applied at each trial location, ranging 0 to 175 lb/ac, achieved from a variety of formulations and application methods. Data on total daily precipitation (in.) and daily maximum and minimum air temperatures (^{o}F) for the days between planting and harvest periods were collected from airport stations located in the areas of interest found in *Weather Underground*. Growing degree days (GDD) represent a "heat value" derived from a comparison of daily maximum and minimum temperatures to a base temperature specific to plant type and provide an estimation of the amount of plant growth achieved at the given time in the growing season (Miller et al., 2001). These values, expressed in degrees Fahrenheit (^{o}F), are calculated as *GDD= (daily max temperature+daily min temperature)/2-base temperature*.and were gathered from *Weather Underground.* Solar radiation data, representing total daily insolation incident on a horizontal surface (MJ/m^{2}/day), were collected from NASA Prediction of Worldwide Energy Resource (POWER) for corresponding latitude and longitude of each experimental site for the respective production season.

Results

Before the estimation, we performed multicollinearity diagnostics (see Table 3 for the correlation matrix). The variables of temperature and growing degree days were highly collinear so we removed the temperature from the estimation. Similarly, the elevation was highly correlated with latitude and longitude; hence the last ones were removed, as well. Evapotranspiration is highly collinear with soil temperature and both variables were not included in the preliminary estimation. Note that our evapotranspiration values are derived from the ET-Penman formula and when there are no available observations, we used the P-Kim computation. These measurement differences may add to the problem of collinearity.

The stochastic production frontier specified in (6) was estimated by the maximum likelihood method assuming half normal distribution of the inefficiency variance. As a robustness check we also estimated the model with least squares and we found an R^2 of 0.60, in that our independent variables explain 60% of the variation of camelina production. The inputs that were statistically significant in the OLS were included as explanatory variables in the stochastic frontier estimation. The average precipitation for the period of interest positively affected the mean camelina yields, whereas it has no impact on yield variability. An increase in average precipitation is correlated with reduced technical inefficiency. Both higher nitrogen rates and higher average growing degree days are correlated with increased average yields. A taller camelina plant is positively correlated with mean yields and yield variability. In contrast, total solar radiation is negatively correlated with mean yields and variation. We also used a non parametric approach to estimate the probability density function of inefficiency. In the Appendix we provide the Kernel density estimates for technical inefficiency scores for the states used in the estimation.

Conclusion

This study develops a model that predicts the probability distribution for expected yields given observed metrics such as precipitation, growing degree days, fertilizer rate, etc. We created a rich database with camelina yields and weather variables from locations in the United States. Applying a stochastic frontier model, we found that precipitation, nitrogen rate, plant height and growing degree days, all positively affected average camelina yields, whereas solar radiation had a negative impact. Future work includes the introduction of additional explanatory variables especially in the estimation of the technical inefficiency. Specifically, we are designing models to test comparative performance of varieties, in conjunction with locations. Tauer and Belbase (1987) found an ambiguous effect between inefficiency and geographic region. We are also looking into variables that will capture the different soil patterns since our data represent a large, diverse area of the US. Soil variables could potentially account for soil interactions with moisture and nutrients. Also, we are investigating whether extremes of water availability and extreme temperatures (e.g., temperature above 32° or below 5°) and varietal characteristics such as drought tolerance exhibit effects on camelina yield distributions. Another future area of study that could help in understanding performance and adoption is the impact of demographics on yield distribution and technical inefficiency, in particular. Furthermore, we will test different distributional assumptions (truncated normal, exponential and gamma) of the efficiency error term.

A limitation of this study is that our data came from multiple sources and account for only one or a few crop seasons so we cannot observe how camelina performed over a longer period of time. We identified multicollinearity as a problem in our analysis so we are working on alternative variable specifications and adding more data to improve descriptive power and fit of the model.

References

AgriLogic Consulting, LLG. "Feasibility Study for Insuring Dedicated Energy Crops-Phase 2: Deliverable: 2.4.2.2." Prepared for Risk Management Agency USDA/RMA/RED, Point of Contact: Joe Davis, 2011.

Aigner, D. J., C. A. K. Lovell, and P. Schmidt. "Formulation and estimation of stochastic frontier production function models." *Journal of Econometrics* 6: 21–37, 1977.

Brandess A., "Modeling the Profitability of Camelina Sativa as a Biofuel Feedstock in Eastern Colorado", *Master's Thesis in Agricultural and Resource Economics, Colorado State University*, Fort Collins, 2012.

Bravo-Ureta, B., D. Solis, V. Moreira-Lopez, J. Maripani, A. Thiam, and T. Rivas. "Technical Efficiency in Farming: A Meta-Regression Analysis." *Journal of Productivity Analysis* Vol. 27, 57-72, 2007.

Burke, I., and P. Fuerst, Washington State University, Dept. of Crop and Soil Sciences, Biofuels Cropping Systems, Cross Cutting Projects, online accessed at: http://css.wsu.edu/biofuels/files/2012/09/CC_Burke_Fuerst_Oilseed_Analysis.pdf

Gesch, R., W., and Archer, D.W., "Double-Cropping with Winter Camelina in the Northern Corn Belt to Produce Fuel and Food", *Industrial Crops and Products,* Vol. 44, p. 718-725, 2013.

Clevenger, W. B., "Evaluating nitrogen rate effect on the camelina variety Cheyenne observed at four dates of planting." Ohio State University Extension, Agronomic Crops Network, 2009, online accessed at: http://agcrops.osu.edu/on-farm-research/research%20reports/2009/Cheyenne%20Camelina.pdf

Colorado State University, The Colorado Agricultural Meteorological Network (CoAgMet), Raw Data Access: online accessed at: http://www.coagmet.colostate.edu/rawdata_form.php

Colorado State University, Extension Soil Crop, Crops Testing, Camelina Results by Year, online accessed at: http://www.extsoilcrop.colostate.edu/CropVar/camelina.html

Crowley, J.G. and A. Fröhlich. "Factors affecting the Composition and Use of Camelina.", Crops Research Center, Oak Park, Carlow, 1998.

Enjalbert, J.N., and J.J. Johnson. "Guide for producing dryland camelina in Eastern Colorado." Colorado State University Extension factsheet No. 0.709, 2011, online accessed at: http://www.ext.colostate.edu/pubs/crops/00709.pdf.

Fullen C., and Polonius, I., "Growing Camelina for Bio-Diesel in Arkansas Delta", Business Feasibility Study for Arkansas Green Energy Network, *Alt. Consulting*, 2011.

Glass, D., “Companies Developing Modified Plant Varieties as Improved Biofuel Feedstocks”, *Blog about Advanced Biotechnology for Biofuels*, 2010. (http://dglassassociates.wordpress.com/category/companies/page/2/)

Guy, S. O., “Camelina: A Potential New Oilseed for Idaho-Agronomic Studies and Cultivar Evaluation.” University of Idaho, Dept. of Plant, Soil and Entomological Sciences, 2007: online accessed at: http://www.extension.uidaho.edu/cereals/nidaho/2007_Camelina_NCI_Crop_Management_Trials.pdf

Jewett, F. G. “Camelina Variety Performance for Yield, Yield Components and Oil Chracteristics.” Ms Thesis submitted to the Colorado State University, 2013.

Johnson, J., N. Enjalbert, J. Schneekloth, A. Helm, R. Malhotra, and D. Coonrod. “Development of Oilseed Crops for Biodiesel Production under Colorado Limited Irrigation Conditions.” Colorado Water Institute, Colorado State University, April 2009 online accessed at: http://water.usgs.gov/wrri/09grants/progress/2009CO186B.pdf

Keske, C.M., Hoag, D.L., Brandess, A., and Johnson, J.J., “Is it economically feasible for farmers to grow their own fuel? A study of Camelina sativa produced in the western United States as an on-farm biofuel”, *Biomass and Bioenergy* 54, 89-99, 2013.

Kumbhakar, S. C., and C. A. K. Lovell. 2000. Stochastic Frontier Analysis. Cambridge: Cambridge University Press.

Lovell, C. A. K. “Production Frontiers and Productive Efficiency.” in H. O. Fried, C. A. K Lovell, S. S. Schmidt (eds.) The Measurement of Productive Efficiency. New York: Oxford University Press, 1993.

Miller, P., Lanier, W., and Brandt, S. 2001. "Using Growing Degree Days to Predict Plant Stages." Montguide. Montana State University Extension Service. online accessed at: http://msuextension.org/publications/AgandNaturalResources/MT200103AG.pdf

Montana State University, Montana Foundation Seed Program, Plant Sciences: online accessed at: http://plantsciences.montana.edu/FoundationSeed/varietyrelease/2007/Camelina.pdf

Montana State University, Northwestern Agricultural Research Center: online accessed at: http://ag.montana.edu/nwarc/research/VarietyEvaluation/VarietyEvaluation.htm#Camelina

Moser, B., R., “Camelina (*Camelina Sativa L.*) oil as a Biofuels Feedstock: Golden Opportunity or False Hope?”, *Lipid Technology*, Vol.22, No.12, 2010.

NASA Prediction of Worldwide Energy Resource (POWER) Climatology Resource for Agroclimatology. 2014. online accessed at: http://power.larc.nasa.gov/cgi-bin/cgiwrap/solar/agro.cgi?email=agroclim@larc.nasa.gov

New Mexico State University Agricultural Science Center, Farmington, 2006 and 2008 Daily Weather Data: online accessed at: http://farmingtonsc.nmsu.edu/weather.html

Newlands, N.K., Townley-Smith, L., and Porcelli, T.A., "A Renewable Source of Jetfuel from Alternative Oilseeds? Predicting Crop Response under Environmental Variability", *Acta Press, A Scientific and Technical Publishing Company*, 2012.

North Dakota State University, Variety Trials, Camelina, online accessed at: http://www.ag.ndsu.edu/varietytrials/camelina

Paulsen, H., M., "Improving Green-House Gas Balances of Organic Farms by the Use of Straight Vegetable Oil from Mixed Cropping as Farm Own Fuel and its Competition to Food Production", *Agriculture and Forestry Research* 3 (61), p. 209-216, 2011.

Pavlista, A. D., and D. D. Baltensperger, "Phenology of Oilseed Crops for Bio-Diesel in the High Plains." *Issues in New Crops and New Uses*, 2007.

Robinson, R. G. "Camelina: A Useful Research Crop and a Potential Oilseed Crop." Minnesota Agricultural Research Station, Station Bulletin 579, 1987.

Roseberg, R. J., and R. A. Bentley. "Growth, Seed Yield, and Oil Production of Spring *Camelina Sativa* in Response to Irrigation Rate and Harvest Method in the Klamath Basin, 2011." Oregon State University Agronomy Research in the Klamath Basin 2011 Annual Report, online accessed at: http://oregonstate.edu/dept/kbrec/sites/default/files/growth_seed_yield_and_oil_production_of_spring_camelina_sativa_in_response_to_irrigation_rate_and_harvest_method_in_the_klamath_basin_2011.pdf

Roseberg, R. J., and R. A. Shuck. "Growth, Seed Yield, and Oil Production of *Camelina Sativa* Grown under Two Irrigation Rate from a Spring Seeding and of Multiple Cultivars Seeded on Multiple Dates in the Fall, in the Klamath Basin, 2007-2008." Oregon State University Agronomy Research in the Klamath Basin 2008 Annual Report, online accessed at: http://oregonstate.edu/dept/kbrec/sites/default/files/2008_spring_and_fall_camelina_final.pdf

South Dakota State University, Dept. of Plant Science, Research, Plant Breeding and Genomics. "Minor Oilseed Research: Camelina." Online accessed at: http://www.sdstate.edu/ps/research/genomics/oilseeds/camelina.cfm

South Dakota State University, West River Ag Center, Crops and Soil Research, Dept. of Plant Science"Annual Progress Report 2009." February 2010: online accessed at: http://www.sdwheat.org/Portals/SDWheat/Special-Projects/WRAC_Crops_Annual_Report_2009.pdf

Sowers, K. E., and W. L. Pan, Washington State University, Dept. of Crop and Soil Sciences "Biofuels Cropping System Project 2009 Final Report." Online accessed at: http://css.wsu.edu/biofuels/files/2012/09/2009_Final_Report.pdf

South Dakota State University Climate and Weather, Automatic Weather Data Network (ADWN): online accessed at: http://climate.sdstate.edu/awdn/soil/graphs.php

Tauer, L. W. and K. P. Belbase. "Technical efficiency of New York Dairy Farms." *Agricultural and Resource Economics Review* 22: 1-9, 1987.

Tindal, C. Renewable Energy Biomass Education Field Day. Knoxville, Tennessee, 2010.

University of Nebraska-Lincoln Extension, Institute of Agriculture and Natural Resource, News from the High Plains Ag Lab, October 2010: online accessed at: http://panhandle.unl.edu/c/document_library/get_file?uuid=5c25152b-a296-4f9d-ac42-005840c2b8d7&groupId=131817

University of Wyoming, Agricultural Experiment Station, Crop Variety Trials: Camelina: online accessed at: http://www.uwyo.edu/uwexpstn/variety-trials/camelina.html

USDA, Natural Resources Conservation Service (NRCS), National Water and Climate Center: online accessed at: http://www.wcc.nrcs.usda.gov/snow/

USDA; NASS. U.S. & All States Data - Crops Planted, Harvested, Yield, Production, Price (MYA), Value of Production, 2010.

U.S. Department of the Interior, Bureau of Reclamation Pacific Northwest Region, AgriMet Cooperative Agricultural Weather Network: online accessed at: http://www.usbr.gov/pn/agrimet/webarcread.html

Washington State University, AgWeatherNet: online accessed at: http://weather.wsu.edu/awn.php?page=dailydata

Weather Underground Weather Forecasts and Reports: http://www.wunderground.com/

Wysocki, D., and N. Sirovatka. "Camelina, a Potential Oilseed Crop for Semiarid Oregon." Oregon State University Extension Service, 2009, online accessed at: http://extension.oregonstate.edu/catalog/html/sr/sr1083-e/sr1083_09.pdf

Young, D.L., Galinato, S.P., and Marsch, T.L., "Feasibility of Camelina as a Biofuels Feedstock in Washington", *IMPACT Center Report, School of Economic Sciences (SES), Washington State University*, 2012.

Zhang, Y., Yu, L., Yung, K-F., Leung, D.Y.C., Sun, F., and Lim, B.M., "Over-Expression of AtPAP2 in *Camelina Sativa* leads to faster Plant Growth and higher Seed Yield", *BioTechnology for Biofuels*, 5:19, 2012.

Appendix

Table 1: Variable Definitions

Variable	Description	Type	Unit
Production Frontier and Technical Inefficiency Effect			
Dependent Variable			
Yield (Y)	natural logarithm of camelina yields	continuous	lbs/acre
Independent Variables			
(conventional inputs)			
Fertilizer (N)	Natural logarithm of nitrogen rare	continuous	lbs/acre
(weather-regional inputs)			
Precipitation (P)	natural logarithm of precipitation from date of planting to date of harvest	continuous	inches
GDD	natural logarithm of growing degree days	continuous	days
Solar (S)	natural logarithm of total solar radiation	continuous	$MJ/m^2/day$
Elevation (E)	natural logarithm of elevation	continuous	feet
Height (H)	natural logarithm of camelina plant height	continuous	inches

Table 2: Summary Statistics of Camelina Yields per state

State (Region)	**Obs**	**Mean**	**Min**	**Max**	**CV***
Colorado	475	982.3	64	2424	58.65
Idaho	27	1568.3	970	2175	26.41
Kansas	42	369.9	59	1370	**87.22**
Minnesota	8	991.1	898.3	1086.5	6.03
Montana	207	1490.7	328.1	2633	39.35
Nebraska	17	712.3	472.8	1025.9	18.89
New Mexico	36	904.8	276.2	1540	40.72
North Dakota	329	1348.5	152	2835	44.22
Ohio	16	978.2	652	1292.7	21.90
Oregon	68	1011.2	61	2278	**64.39**
South Dakota	50	1046.2	494	1476	26.36
Washington	98	1705.4	1024	3379	23.76
Wyoming	173	762.4	96	2041	**68.87**

*** Coefficient of Variation**

Note: State values are not directly comparable due to inconsistent trial conditions and highly heteroskedastic trial data.

Table 3: Camelina Yields of the "most common" varieties

Variety	**Obs**	**Mean**	**Min**	**Max**	**Variance**	**Skewness**
Calena	96	1406.8	104	3089	483607.7	0.047
Ligena	81	1394.5	125	3023	427105.5	0.130
Blaine Creek	66	1253.6	64	2519	416894	0.123
Suneson	63	1261.8	79	2298	350462.3	-0.185
Celine	62	1068.8	117	3379	386333.3	0.806
Cheyenne	72	1073.07	123	2275	306971.7	0.139
Robinson	41	1276.9	204	2481	405806	0.150
Galena	23	1508.3	407	2667.3	513491.9	0.218
Yellowstone	21	1001.9	94	1911	483656.4	0.244
BSX G22	18	939.5	95	1798	352377.9	0.004
Lindo	18	947.1	80	1862	404267	0.116
Licalla	17	895.5	120	1883	348599	0.176
BSX G24	16	906.3	103	1809	386830.9	0.249

Note: Variety values are not directly comparable due to inconsistent trial conditions and highly heteroskedastic trial data.

Table 4: Correlation Matrix

	Soil Temp.	Evapotransp	Nitrog. rate	Seed rate	Land	Air temper	Precip	GDD	Wind speed	Elev	Latit	Longit	Solar Rad.
Soil Temp													
Evapotransp	**0.9872**												
Nitrog. rate	0.8683	**0.9067**											
Seed rate	-0.6802	-0.7452	-0.7671										
Land	0.8421	0.7708	0.4651	-0.3650									
Air temper	0.6966	0.7665	**0.9595**	-0.7465	0.1997								
Precip.	-0.7641	-0.7091	-0.3514	0.3391	**-0.9733**	-0.0907							
GDD	0.4959	0.5726	0.8611	-0.6372	-0.0464	**0.9642**	0.1714						
Wind speed	-0.8458	-0.7753	-0.4712	0.3700	**-1.0000**	-0.2065	**0.9728**	0.0396					
Elev.	0.5626	0.6601	0.8867	-0.7287	0.0321	**0.9777**	0.0476	**0.9701**	-0.0391				
Latit.	-0.6375	-0.7276	**-0.9238**	0.7576	-0.1258	**-0.9892**	0.0443	**-0.9585**	0.1328	**-0.9956**			
Longit.	0.4075	0.2931	-0.0924	0.0906	0.8345	-0.3688	-0.8557	-0.5743	-0.8307	-0.5239	0.441		
Solar Rad.	**0.9761**	**0.9973**	**0.9328**	-0.7659	0.7224	0.8107	-0.6565	0.6293	-0.7272	0.7133	-0.775	0.2224	

Table 5: Parameter Estimates from Stochastic Frontier Regression (half normal distribution) ⁎ (N=165)

	Mean	**Random Error variance (υhat)**	**Inefficiency variance (uhat)**
Variable	Coefficient (st.error)	Coefficient (st.error)	Coefficient (st.error)
Constant	12.445 *** (2.589)	2.106 (20.715)	2.106 (20.715)
Precipitation (P)	0.409 *** (0.100)	1.662 (4.653)	-1.211 ** (0.451)
Fertilizer (N)	1.251 *** (0.130)	N/A	N/A
Height (H)	0.352 ** (0.177)	15.705 ** (6.594)	-0.467 (1.399)
GDD	1.702 *** (0.216)	-4.038 (4.156)	-0.356 (0.739)
Solar Radiation (S)	-1.689 *** (0.184)	-9.230 * (5.192)	-0.525 (0.587)
Elevation (E)	0.140 (0.489)	35.628 (45.304)	-0.010 (2.520)

Notes:
⁎ allows for heteroskedastic error terms
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 6: Parameter Estimates from Least Squares Regression (N=165)

Variable	Coefficient (st.error)
Constant	7.405 (6.284)
Precipitation (P)	0.750 ** (0.236)
Fertilizer (N)	1.393 *** (0.246)
Height (H)	0.595 * (0.312)
GDD	1.801 *** (0.165)
Solar Radiation (S)	-1.764 *** (0.325)
Elevation (E)	0.709 (0.899)
R^2=0.6085	

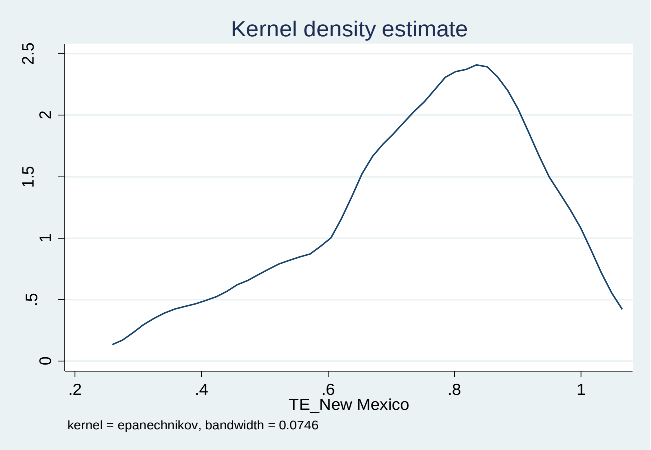

Figure 1: Kernel Density Estimates for Technical Inefficiency Scores, New Mexico

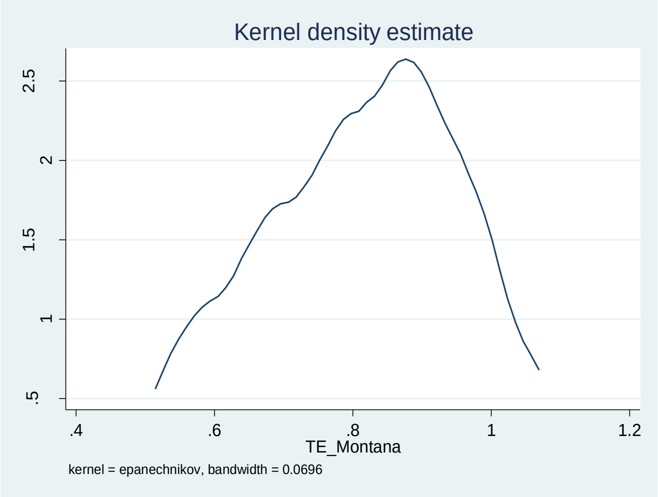

Figure 2: Kernel Density Estimates for Technical Inefficiency Scores, Montana

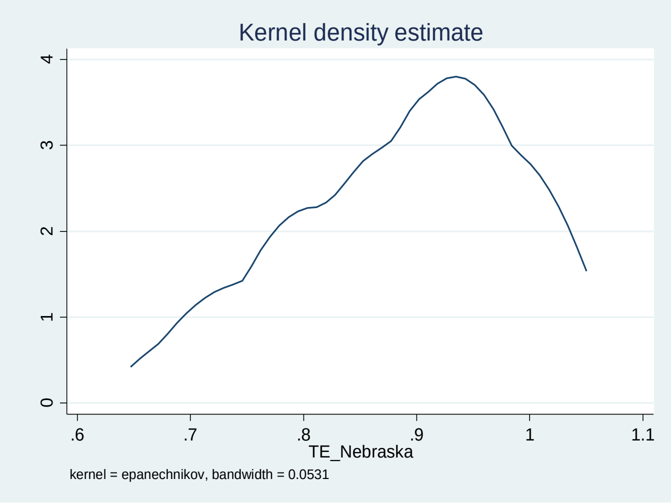

Figure 3: Kernel Density Estimates for Technical Inefficiency Scores, Nebraska

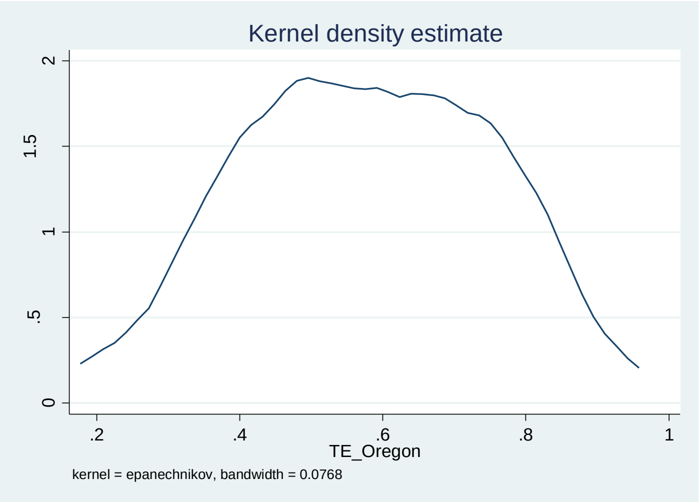

Figure 4: Kernel Density Estimates for Technical Inefficiency Scores, Oregon

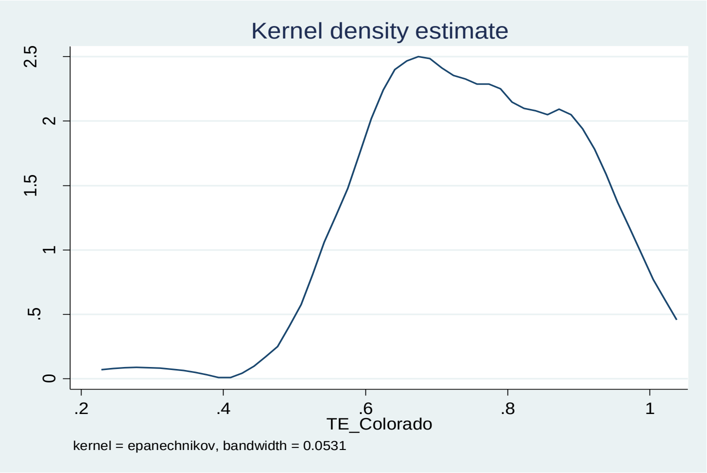

Figure 5: Kernel Density Estimates for Technical Inefficiency Scores, Colorado