Application of Linear and Non-linear Models into Trend Analysis of U.S. Cotton Export (1996-2017)

Zahra Saki, Marguerite Moore, Lori Rothenberg
Department of Textile and Apparel, Technology and Management, Wilson College of Textiles, NC State University, Raleigh, NC 27606

Abstract
Cotton fiber is the largest source of U.S. textile and apparel export advantage. Rises and falls in cotton export advantage over the examination period (1996-2017) embolden research into past trend analysis of cotton fiber. UN-COMTRADE database is used to extract export values of products including HS5201, *Cotton; not carded or combed*, to calculate the *Normalized Revealed Comparative Advantage* (NRCA) index. Simple linear regression (SLR), modified SLR following outlier removal and a non-linear model (i.e., cyclical) are used to fit the data. A cyclical model suggesting a repeated period of seven years best describes the variation of U.S. cotton export competitiveness.

Key Words: Trend analysis, Linear models, Non-linear model, Textiles and Apparel, Comparative Advantage, Cotton

Data and Measure

The global trade of cotton fiber for the most recent 22-year period is selected from the United Nations Commodity Trade Database (UN COMTRADE). To evaluate the past trend of U.S. cotton comparative advantage, Normalized Revealed Comparative Advantage (NRCA) values are calculated for the period between 1996-2017 and are fit using linear and non-linear models. NRCA (Yu, Cai, and Leung, 2009) is a metric that measures export comparative advantage of a country in a given product and is defined as:

$$NRCA = (E_j^i/E) - (E_j * E^i/E * E) \quad (1)$$

Where E_j^i stands for the country i export of commodity j,
E_j refers to the export of commodity j by all countries in the world,
E^i is the country i export of all commodities,
and E is the export of all commodities by all countries.
NRCA is symmetric around zero and values greater than zero indicate a country's comparative advantage in a given commodity.

Results

A simple linear regression model (equation 2) is used as the simplest model to fit the data.

$$NRCA_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (2)$$

Declines in NRCA value in the year 1999, encourages the removal of this datapoint and subsequent refit of the simple linear regression model. This decline in NRCA value is likely associated with the 1998 U.S. drought (USDA, 1998) and is therefore considered as an

outlier observation. Simple linear and modified simple linear regression fits are demonstrated in Figure One.

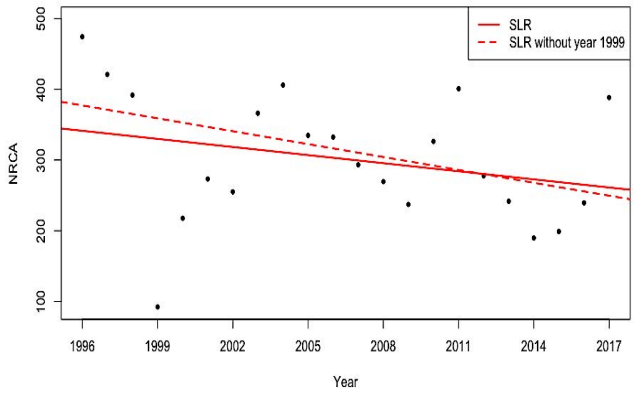

Figure 1: U.S. cotton NRCA, overlaid with simple linear regression models with and without the year 1999

Though the adjusted-R^2 improved after removing the year 1999 (Table 1), the observed residual pattern versus fitted value graph suggests missing terms in the structural portion of the model. Therefore, polynomial models are applied to improve the fit.

Table 1: Results of linear models

	SLR	SLR*
R-squared	0.072	0.23
Adj. R-squared	0.026	0.19
F-statistic	1.56 (1,20)**	5.73 (1,19)**
P-value	0.22	0.03

* Without year 1999
**Number in parenthesis are Degrees of freedom

Using subset selection and (Bayesian Information Criterion) BIC (James et al., 2013), the model in equation three was selected to fit the data.

$$NRCA_t = \beta_0 + \beta_1 t + \beta_2 t^4 + \beta_3 t^7 + \beta_4 t^8 + \varepsilon_t \quad (3)$$

The polynomial model did not improve the fit and introduced a problem in terms of model structure as it does not include all degrees of polynomials smaller than eight. Rises and falls in NRCA values over time suggest that a sinusoidal function with regression parameters that define the location (phase) and height (amplitude) of the peak is a reasonable candidate to fit the data. For the cyclical model both periods seven and eight are fit to the data (Figure 2), and the result suggests that the model with a period of seven indicates an improved adjusted R^2 (Table 2).

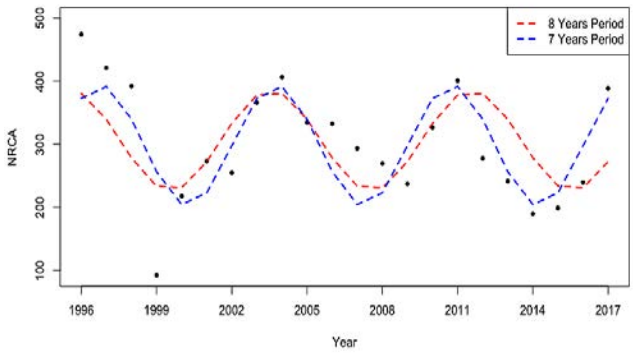

Figure 2: U.S. cotton NRCA overlaid with cyclical models for periods seven and eight

Table 2: Results of cyclical models

	Cyclical period 8	Cyclical period 7
R-squared	0.39	0.55
Adj. R-squared	0.33	0.53
F-statistic	6.13	11.17
P-value	0.0088	0.0007

Estimated parameters for the cyclical model, for period seven, are presented in equation four.

$$NRCA = 297.87 + 95.77Sine(7) - 0.3\ Cosine(7) \quad (4)$$

At this point, there is no clear explanation for the observed pattern that export comparative advantage of U.S. cotton appeared to cycle every seven years among the most recent 22 years. To address this finding, future research incorporating observations before 1996 as well as qualitative research with direct expertise in this field will likely provide insight into this observation.

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Saki, Z., Moore, M., Kandilov, I., Rothenberg, L. & Godfrey., A. B. (2019). Revealed Comparative Advantage for U.S. Textiles and Apparel. *Competitiveness Review: An International Journal,* Vol. 29 Issue: 4.

USDA (1998), "Outlook for US agricultural exports ", available at http://usda.mannlib.cornell.edu/usda/ers/AES//1990s/1998/AES-12-03-1998.asc

Yu, R., Cai, J. and Leung, P. (2009), "The normalized revealed comparative advantage index", *The Annals of Regional Science*, Vol. 43 No. 1, pp. 267-282.