Semi-Parametric Generalized Additive Vector Autoregressive Models of Spatial Basis Dynamics

Selin Guney, Barry K. Goodwin, and Andrés Riquelme

An extensive line of research has examined linkages among spatially-distinct markets. We apply semi-parametric, generalized additive vector autoregressive models to a consideration of basis linkages among North Carolina corn and soybean markets. An extensive suite of linearity tests suggests that basis and price relationships are nonlinear. Marginal effects, transmission elasticities, and generalized impulse responses are utilized to describe patterns of adjustment among markets. The semi-parametric models are compared to standard threshold vector autoregressive models and are found to reveal more statistical significance and substantially more nonlinearity in basis adjustments. Marginal effects are nonlinear and impulse responses suggest greater adjustments to extreme shocks and asymmetric adjustment patterns. The results provide evidence in favor of efficiently linked markets.

Key words: Law of one price, generalized additive models, nonlinear time series models.

JEL codes: C58, Q02.

The empirical literature addressing the behavior of commodity prices over time and across spatially-distinct markets has grown substantially and remains an important research question. A fundamental axiom of economics—the "Law of One Price"—underlies the arbitrage behavior thought to discipline and characterize such relationships. This literature has progressed from a simple consideration of correlation coefficients and linear regression models to classes of models that address particular time-series properties of price data and consider nonlinear price linkages.

In many cases, it is reasonable to assume that commodity prices act in a way that is consistent with stable behavior around a shifting or breaking mean. In recent years, the literature analyzing this phenomenon has focused on models capable of accommodating regime change and mean-shifting behavior in order to determine the nature and character of price relationships, and of changes in prices over time and across different markets. This mean-shifting behavior and the reasons behind these price movements have been addressed through the application of various specifications and econometric techniques (see, e.g., Ng and Voselgang 2002; Enders and Holt 2012; Enders and Holt 2014b; Holt and Teräsvirta 2012).

Enders and Holt (2012) applied nonstructural time series models, namely standard vector autoregressive models (VAR) and a more flexible shifting mean VAR (SM-VAR) approach, which allowed for smoothly-shifting means. These authors incorporated the methods developed by Bai and Perron (1998, 2003), Becker, Enders and Hurn (2004), Becker, Enders, and Lee (2006), and Gonzalez and Teräsvirta (2008) to capture the timing and to describe the nature of structural breaks (shifts) in price relationships for a variety of commodity markets.

This literature has evolved to consider new methods for statistically testing structural

The helpful comments of three anonymous referees are gratefully acknowledged. Particular thanks are due to Thomas Yee, Jim Vercammen, Stephan von Cramon-Taubadel, and Zheng Li for their invaluable suggestions. Guney is an assistant professor in the Department of Agriculture at Texas State University. Goodwin is William Neal Reynolds Distinguished Professor in the Departments of Economics and Agricultural and Resource Economics at North Carolina State University. Riquelme is an assistant professor on the Facultad de Economia y Negocios of the Universidad de Talca, Chile. Goodwin's research was supported by the NC Agricultural Research Service, the U.S. Forest Service, and BARD Grant No. IS-4693-14. Correspondence to be sent to: barry_goodwin@ncsu.edu.

Amer. J. Agr. Econ. 101(2): 541–562; doi: 10.1093/ajae/aay033
Published online June 27, 2018

change and mean-shifting behaviors. Chow tests with known break points have evolved into tests of discrete and gradual conditional mean shifting with unknown break points and variable speeds of adjustment among regimes. These tests address the widely-recognized problems associated with non-standard test statistics and parameters that may be unidentified under null hypotheses. A variety of alternative tests for (non-) linearity have been developed and are often used to detect departures from linearity in empirical models.

Linkages among spatially separated markets for a homogeneous commodity are often examined in an attempt to characterize relationships among such markets and the overall functioning of an aggregate market. As Fackler and Goodwin (2001) note, the strength of such linkages and the extent to which isolated shocks are transmitted to other spatially-distinct markets remains an important indicator of market performance and "efficiency." If spatially separated markets exist in isolation and do not respond to aggregate or localized shocks, barriers to trade such as poor infrastructure, trade-inhibiting government policies, and limited transfer of market information may exist.[1] These efficiency issues are often a particular concern in developing economies, where poor infrastructure may inhibit spatial trade. However, evaluation of market integration is often used to investigate performance and behavior in all types of markets for spatially-traded goods.

This article proposes a new class of semi-parametric time series models that accommodate nonlinear price adjustment behavior of an a priori unknown type in a vector autoregressive modeling framework. This approach is viewed as a natural next step in the evolution of nonlinear time-series models of spatial and regional price behavior. To this end, recent advances in semi-parametric modeling that have developed methods for additive models consisting of a mixture of parametric and nonparametric components are considered. These vector autoregressive models adopt the Generalized Additive Models (GAM) specification and estimation procedures proposed by Hastie and Tibshirani (1986), Yee and Wild (1996), Linton (2000), and Yee and Hastie (2003). In particular, the penalized, iteratively-reweighted maximum likelihood algorithms developed for GAM model estimation by Wood (2004) are used to estimate a fully flexible semi-parametric model of spatial price relationships. The estimates reveal significant evidence of nonlinear patterns of adjustment and nonlinear trend effects.

The application in this paper is to daily price data collected from a number of important, spatially-distinct North Carolina corn and soybean markets. Versions of these price data have been previously utilized to evaluate regional price linkages and spatial market integration (see, e.g., Goodwin and Piggott 2001, Sephton 2003, and Greb et al. 2013). Nonparametric marginal effects and impulse response functions are used to evaluate the dynamics of regional price adjustments to localized shocks in individual markets. Implications for regional price adjustments and, in particular, adjustments during periods of large swings in prices, are discussed. We depart from existing research on this topic in one important way. We consider price linkages among basis series in local markets. Prices are adjusted by subtracting the (logarithmic) futures prices from each (logarithmic) local price before comparisons are made. This is done to address concerns that spatial linkages may be heavily dependent upon movement in the aggregate market.

Several characteristics of the agricultural industry in North Carolina make this application particularly interesting. According to the 2012 *Agricultural Census*, North Carolina is the second-largest producer of hogs and turkeys, and the fourth-largest producer of broilers and chickens among U.S. states. However, the state is only the eighteenth-largest producer of corn and the fifteenth-largest producer of soybeans. The implication is that the state typically has a huge deficit in corn and soybeans and is very dependent upon a well-functioning transportation system to distribute grain from growing areas and within the state. The specific markets considered here are spatially dispersed, ranging from 86 to 277 highway miles apart.[2]

[1] An example pertinent to the markets considered here arises in the North Carolina General Statutes Sec. 20-118, which places significant restrictions on grain transporters operating over distances greater than 150 miles, but waives such restrictions for local transportation.

[2] Basis relationships for local soybean markets in North Carolina are discussed by Piggott et al. (2017). These authors note that local basis may be positive or negative, depending on

The Law of One Price (LOP) has been extensively investigated because it has important implications both for economists and traders. The law implies that no persistent opportunities for profitable spatial arbitrage exist. It may also have policy implications in that it provides an empirical measure of market interactions that may help policymakers to determine and evaluate regional and international trade policies. The general conclusion underlying this broad conjecture is that prices for homogenous products at different geographical locations should not differ more than transport and transaction costs, which include such things as freight costs, insurance, and contract fees. An obvious reason why the prices of homogenous products may not be the same in different locations is due to the aforementioned transaction and transport costs and other impediments to trade such as tariffs and quotas. As a result, deviations from the LOP may contain significant nonlinearities and may exhibit persistent departures from parity conditions. Transactions costs are typically unobservable (at least partially) but nonetheless may impact spatial trade and price behavior. The presence of such transactions costs has typically been represented using nonlinear models, where nonlinearities reflect the fact that price adjustment patterns may be different for large shocks than for small ones. Specifically, shocks large enough to imply profitable spatial arbitrage may evoke significant responses in regional markets whereas small shocks that are not significantly larger than transactions costs may not evoke equilibrating price responses.

Most recently, following these theoretical arguments, several studies have employed nonlinear models to investigate the validity of the LOP. Among these studies are those by Michael, Nobay, and Peel (1994), Obstfeld and Taylor (1997), Taylor (2001), and O'Connell and Wei (2002). In these studies, the nonlinear nature of the adjustment process is generally investigated in terms of some version of a threshold autoregressive (TAR) model. The studies provide cumulative evidence in favor of the threshold-type nonlinearity in deviations from the LOP.

Among the studies that use variants of discrete threshold models are Balke and Fomby (1997), Goodwin and Piggott (2001), Lo and Zivot (2001), Sephton (2003), and Park, Mjelde, and Bessler (2007), all of whom have found support for the validity of LOP and the presence of threshold effects. Additionally, these studies conclude that the path of adjustment to equilibrium often depends on the size of the shock introduced into the system. However, reasons exist to think that the patterns of price adjustment in the markets are smooth rather than discrete, despite the discrete nature of the observed economic behavior underlying the adjustments, that is, arbitrage is either profitable or not, (Goodwin, Holt, and Prestemon 2011). This smoothness reflects the fact that market prices typically represent the aggregated actions of a number of individual traders. The literature has progressed to reflect the use of smooth transition models instead of discrete models of transition (see, e.g., Enders and Holt 2012, and Goodwin, Holt, and Prestemon 2011). The semi-parametric models considered in this paper admit various forms of regime changes, which may occur gradually or at discrete points in time.

Econometric Methods

Our empirical approach considers two related methodologies. Following a well-established and voluminous literature, we consider discrete threshold models. Such models have been widely applied to considerations of spatial price linkages, including applications to some of the same markets considered here (e.g., Goodwin and Piggott 2001, Sephton 2003, and Greb et al. 2013). We evaluate both bivariate and trivariate analyses of spatially separate markets. We then consider fully nonlinear, semi-parametric models.

In contrast to existing research, we consider local basis series (defined as the difference between [logarithmic] local prices and contemporaneous prices on the CBOT/CME) rather than just local prices.[3] This has two important implications for our analysis. First, one would expect basis relationships to be stationary. If not, the prices could drift

the time of year. For example, they show that soybean basis in Fayetteville, North Carolina, is weakest (and negative) in the fall and strongest in the late summer.

[3] We are grateful to an editor for suggesting this approach.

arbitrarily far apart in the long run. A second point important to our threshold modeling is implied by the results of stationarity tests. Namely, if the bases are stationary, a threshold vector autoregressive (TVAR) model is appropriate rather than an error-correction model expressed in differences. Existing research has typically considered threshold error-correction models in light of the typical finding of nonstationary commodity prices. A vector error correction model (VECM) is equivalent to a restricted version of a standard vector autoregressive model in levels of the nonstationary data. Kilian and Lütkepohl (2017) discuss the issues associated with this distinction and note that whether one prefers the VECM approach or the VAR in levels when considering cointegrated, nonstationary variables comes down to how certain one is of the implied restrictions. These authors also note that when the model specification is guided purely by statistical tests (such as for cointegration), the results have to be interpreted with caution. Simulation evidence in Gospodinov, Herrera, and Pesavento (2013) suggests that estimates based on the VAR model in levels are typically more accurate than estimates from models selected on the basis of pretesting.

The next step in our analysis involves a consideration of a suite of tests intended to detect departures from linearity in conventional time-series models. A variety of (non-) linearity tests were conducted for the price data. A standard "self-exciting" threshold autoregressive (SETAR) model of the form applied to prices in spatially-distinct markets by Goodwin and Piggott (2001) was considered for each of the market pairs. This specification is given by:

$$(1) \quad \Delta y_t = (\lambda_0^1 + \lambda_1^1 y_{t-1}) I_A(y_{t-1} \le c_i) + (1 - I_A(y_{t-1} \le c_i))(\lambda_0^2 + \lambda_1^2 y_{t-1})$$

where y_t is the differential between two spatially distinct prices, $(y_t = p_t^1 - p_t^2)$, $I_A(\cdot)$ is the indicator function, c_i is a threshold value, and (λ_i^j) is the jth parameter for regime i. The test for nonlinearity involves the null hypothesis that $\lambda_k^i = \lambda_k^j$ for $(k = 1, \ldots, p)$. Because alternative regime parameters are undefined under the null hypothesis, the test statistic is nonstandard. We utilize Hansen's modification of standard Chow-type tests for each pair of markets.[4] We also applied Tsay's (1989) linearity test to the price differentials. Tsay's test orders the data according to the value of the threshold variable (the lagged price differential in this case) and calculates out of sample recursive residuals. An F-test of linearity is given by the regression F statistic derived from regressing the recursive residuals on the ordered threshold variables.

We also applied two versions of neural network tests of linearity. The linearity test of Teräsvirta, Lin, and Granger (1993) is based upon a Taylor-series expansion of a neural network model with a single hidden layer. Linearity is implied if the optimal weights of the network are zero. The neural network is defined using logistic activation or "squashing" functions.[5] White's (1989) test is identical except that it tests the activation functions directly without the series expansion.

Finally, we considered multivariate threshold VAR models containing all three prices. In this case, we utilized the lagged residual from a regression of one price (Candor for corn and Fayetteville for soybeans) on the other two prices as a forcing variable (i.e., the variable used to identify the thresholds). Hansen's multivariate test is applied to the 3-variable threshold VAR (TVAR) model. This specification is analogous to the vector error correction (VECM) models often applied to spatial price models. In the SETAR and TVAR threshold models, we consider specifications with both one and two thresholds.

Nonparametric regression allows the assumption of linearity to be relaxed, which may be proper for many economic relationships, thus allowing a visual exploration of the data to uncover structure that might otherwise be missed when evaluated in a parametric form. It is widely recognized that many forms of nonparametric regression do not work well when the number of independent variables in the model is large. In such cases, a large data set is needed to avoid the "curse of dimensionality," which is defined as the problem of rapidly increasing variance as the number of parameters increases (i.e., as the flexibility of a specification is increased).

[4] All of the bootstrapping results presented in this paper utilized 500 replications.

[5] See Hornik, Stinchcombe, and White (1989) for a detailed discussion of neural networks and activation functions.

Another pitfall of using nonparametric regression is that the results and the implications for economic relationships among variables may sometimes be hard to interpret.

To overcome these problems, Stone (1985) proposed additive models that adopt an additive approximation to the multivariate regression function. By doing so, the curse of dimensionality problem is overcome because each individual additive term is estimated using a univariate smoother separately, and the approximation is obtained locally rather than universally. The aforementioned interpretation problem is avoided as the estimates of the individual terms explain how the dependent variable changes with the independent variables.

Extensions of additive semi-parametric models that are valid for a wide range of distributional families such as the exponential have been proposed by Hastie and Tibshirani (1986) through the application of Generalized Additive Models (GAM) that enable the mean of the dependent variable to depend on an additive predictor through a linear or nonlinear link function. The basic GAM modeling framework used to investigate the basis relationships may be stated as follows. Let y be a response random variable (prices for example) and $x_1, x_2, \ldots x_p$ be a set of predictor variables (such as related basis and lagged values). A regression procedure can be viewed as a method for estimating the expected value of y conditional on the values of $x_1, x_2, \ldots x_p$. The standard linear regression model assumes a linear form for the conditional expectation:

$$(2) \quad E(y|x_1, x_2, x_p) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p.$$

The additive model generalizes the linear model by modeling the conditional expectation as a sum of smooth functions of the covariates. Wood (2004) presents the following model:

$$(3) \quad E(y) \equiv \mu, \qquad \text{and}$$

$$(4) \quad g(\mu) \equiv \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + s_1(x_{k+1}) + \ldots + s_p(x_p) + \epsilon_i$$

where ϵ_i follows a specific parametric distribution (Gaussian in our case), $g(\cdot)$ is a monotonic link function (linear in our case), and $s(\cdot)$ represents smoothed nonparametric functions of the covariates. Specific models may contain the smoothed functions alone or may include combinations of linear and smoothed covariate terms. The smoothing functions $s(\cdot)$ are nonparametric and are estimated using one of a variety of nonparametric methods, such as local polynomials, running mean and median smoothers, nearest neighbors, or splines. Additive models have the advantage of avoiding overfitting since all smoothing is done nonparametrically. The additive framework also has an advantage, as noted by Yee (2015), of being straightforward to interpret. Each effect can be evaluated while holding all other covariates constant. The cost of such simplicity is that interactions among covariates are less easily represented.

If a parametric distribution is specified for the distribution of ϵ_i, it is straightforward to define a likelihood function and proceed with estimation methods that are analogous to maximum likelihood (as well as standard regression).[6] Yee and Wild (1996) discuss iterative reweighting estimation procedures for vector generalized additive models. We utilize thin plate regression splines with penalized higher-order derivatives of the basis functions applied to a set of 2,000 equally-spaced knot points. Thin plate splines are represented as a sum of radial basis functions, the parameters (a subset of α) of which are estimated using penalized regression methods. The radial basis functions are kernels of the form

$$(5) \quad s(x) = \sum_{i=1}^{K} \alpha_i \varphi(||x - x_0||)$$

where $\varphi(||x - x_0||) = ||x - x_0||^2 \log(||x - x_0||)$, x_0 is a reference point, $||\cdot||^2$ is the Euclidean norm, and α_i terms represent basis function parameters to be estimated.[7] A nonparametric additive model can be estimated by

[6] Yee (2015) notes that this approach is only "quasi-maximum likelihood" as the error term does not have a precise parametric definition in nonparametric models. Estimation methods estimate parameters under the assumption that the likelihood function depends only on the first two moments.

[7] Detailed discussions of thin-plate splines are contained in Bates et al. (1987), Wahba (1990), and Wood (2003). The use of thin plate regression splines rather than thin plate splines, which essentially have knots at every point, is made to make nonlinear estimation possible. A thin plate spline applied to a data set of size n will have $n - 1$ basis functions, making estimation very complex.

standard ordinary least squares or maximum likelihood methods by choosing the basis function parameters that minimize $||y - s(x)||^2$, where $s(x_p)$ is a vector of splines.[8] The shortcoming of this approach is that there is nothing to control the smoothness of $s(x)$. To this end, a penalty is applied and the basis functions are estimated by minimizing

$$(6) \quad ||y - s(x)||^2 - \lambda J_m(s)$$

where J_m is a penalty comprised of derivatives m and higher of the $s(x)$ functions.[9] Our application limits the penalty to second derivatives of the spline. In the case of multiple (p) covariates in an additive model and thus multiple one-dimensional splines arranged in an additive form, a p-dimensional set of λ_i smoothing parameters must be specified.

If we define $L(\alpha)$ to be the log of the likelihood function for a given parametric distribution, conditional on a set of given smoothing terms λ, we can write the penalized maximum likelihood estimation (MLE) problem in a form entirely analogous to least squares regression:

$$(7) \quad L_p(\alpha) = L(\alpha) - \frac{1}{2}\alpha' J_\lambda \alpha$$

where $L(\alpha)$ represents the standard log likelihood function without penalty, $L(\alpha)' - J_\lambda\alpha$ is the gradient of the likelihood function, and $L''(\alpha) - J_\lambda$ is the corresponding Hessian matrix. In an approach entirely analogous to ordinary least squares, the parameters α can be estimated as

$$(8) \quad \hat{\alpha} = (X'X + J_\lambda)^{-1}X'y$$

which implies an influence matrix of $H_\lambda = (X'X + J_\lambda)^{-1}X'$, which is equivalent to the OLS version with the addition of the penalty matrix.

The influence matrix plays an important role in developing a metric for use in estimating the smoothing terms λ, which have until now been assumed to be given, as well as providing a measure of the degree of nonlinearity inherent in the response. The dimension of the basis functions must be truncated at some point. We allow for a maximum rank of 9 for J_λ.[10] A variety of methods can be used to estimate smoothing terms. We consider an out-of-sample generalized cross-validation (GCV) approach. The GCV model evaluation criterion chooses the λ that minimizes

$$(9) \quad GCV_\lambda = \frac{n\sum_{i=1}^{n}(y_i - \hat{s}(x))^2}{(tr||(I - H_\lambda)||)^2}$$

where tr is the trace of the matrix. The numerator represents the squared difference between actual and observed values of y. Once basis parameters and smoothing terms are estimated, the effective degrees of freedom for the entire model is given by the trace of the influence matrix (trH_λ). Each spline term has an effective degrees of freedom given by the trace of the submatrix of the influence matrix that corresponds to each particular spline. Thus, $DF_j = tr(H_{j\lambda})$ and the sum of the degrees of freedom for each component is equal to the trace of the entire influence matrix, which is equivalent to the sum of individual component degrees of freedom, and the degrees of freedom corresponding to parametric estimates (an intercept and dispersion parameter in our application). The derivative-based penalties can also be summed across individual splines to yield an overall roughness penalty. The dispersion parameter in our application of a normal likelihood function corresponds to the estimated model standard error σ. Each equation of the model also contains an additive intercept term. A nonlinear response is implied for values of the effective degrees of freedom parameter that are greater than one. Nychka (1988) has shown that a conventional Bayesian covariance matrix of the form

$$(10) \quad V_\alpha = (X'X + J_\lambda)^{-1}\sigma^2$$

can be used to derive confidence intervals and significance tests that are consistent with a frequentist interpretation. This yields a

[8] A point of clarification regarding nomenclature is helpful. In reference to market prices, basis is the difference in a local price and a futures price. In the terminology of splines, basis functions are the individual piecewise polynomials that make up the splines. Despite the common terminology, the concepts are distinct.

[9] Thin-plate splines may include multiple covariates. We restrict the additive model to one-dimensional splines. In this case, the penalty for each $s(x_i)$ term is $J_i = \int_{-\infty}^{\infty} (\partial^2 s(x_i)/\partial x_i^2)^2 dx_i$.

[10] The empirical results presented below were not sensitive to allowing a higher dimension basis.

direct test of statistical significance of each individual effect.

As noted, interpretation of the results depends on the effective degrees of freedom parameters (which indicate the implicit nonlinearity in the responses). We also consider the estimated marginal effects, which vary across the distribution of each variable, and derivatives of the marginal effects ($\partial s(x_i)/\partial x_i$), which indicate the effect of changes in a covariate (a lagged logarithmic price) on the dependent variable (a current logarithmic price).[11] The aforementioned presence of unobservable transactions costs suggests that these derivatives should suggest greater response to bigger price changes. In particular, the price transmission elasticities should reflect greater adjustment when price shocks are large. In comparison, price elasticity estimates in a standard linear (in logarithms) VAR model will yield a constant elasticity. In cases where the effective degrees of freedom parameter approaches one, the response is equivalent to a linear model and a constant elasticity is implied.

For each spline, the GCV metric is used to identify the optimal smoothing parameters and thus the effective degrees of freedom. The smaller is the smoothing parameter, the less smoothing that is applied and thus the more nonlinear will be the spline function and the larger will be the effective degrees of freedom. Likewise, a smaller roughness penalty corresponds to less of a penalty, a higher effective degrees of freedom, and a smaller smoothing parameter.

In addition, we pursue generalized impulse response functions. With such nonlinear models, the nature of such responses may depend upon the timing, size, and direction of the shock. We consider 1% positive and negative shocks to each basis series and evaluate the impulses. A generalized impulse is given by:

$$(11) \quad GIF(Y_{it+j}) = E(Y_{t+k}|Y_t + \eta, \ldots Y_{t-j}) - E(Y_{t+k}|Y_t, \ldots Y_{t-j})$$

where $\eta = +/-.01$ in our application. As Koop, Pesaran, and Potter (1996) note, a variety of different generalizations of standard impulse responses can be considered for nonlinear models. Following Goodwin and Piggott (2001), we evaluate the impulse responses at the last observation. This is necessitated by the very long computational time required for each replication of the model.[12]

To summarize, we apply generalized additive modeling techniques to generate a fully nonparametric (in additive form) vector autoregressive model of local basis for corn and soybeans. To our knowledge, this is the first application of such nonparametric modeling techniques to a general model of spatial price behavior. This application provides insights into spatial linkages among an important set of regional agricultural markets and adds to the growing body of research applying nonlinear estimation techniques to models of spatial price parity. We also estimate threshold vector autoregressive (TVAR) models in the spirit of Goodwin and Piggott (2001), Sephton (2003), and Greb et al. (2013), who applied TVAR models to allow for nonlinearities among similar prices for corn and soybeans in North Carolina. The TVAR estimates and specification testing results provide a benchmark against which the VGAM model results can be compared. Price responses to changes in other prices are interpreted through marginal effects, nonparametric price transmission elasticities, and generalized impulse responses.

Data and Empirical Application

Our application is to daily corn and soybean prices observed at three commercially important North Carolina markets. Corn prices were obtained at Candor, Cofield, and Roaring River, whereas the prices for soybeans were quoted at Fayetteville, Cofield, and Norwood. The price data span the period of January 1, 1990 to February 23, 2017. On holidays where all prices were missing in each of the markets mentioned, the observations were omitted from the sample and a smooth

[11] The derivatives represent basis transmission elasticities ($\varepsilon_{ij} = \partial \log(p_t^i)/\partial \log(p_{t-s}^j)$), which are allowed to vary in a nonparametric fashion across the span of each covariate.

[12] Confidence intervals on the VGAM impulses are calculated using 500 replications of size n (n=6,673 for corn and 6,436 for soybeans) with replacement. Impulses for each replication are generated and used to define a 90% confidence interval for the responses. For the TVAR model, the covariance matrix of the parameter estimates, conditional on the thresholds, is used to generate Monte Carlo simulation estimates of the impulses. In each case, prices with and without a shock are dynamically forecasted 48 days into the future and the impulse is given by the difference in forecasts.

continuity of the prices was assumed. In cases where individual prices were missing, cubic spline interpolation was used to replace missing values.[13] Logarithmic transformations of the basis series are the focus of the empirical analysis, and the aforementioned cities were chosen on the basis of market prominence and data availability. Data availability largely reflects the prominence and durability of individual markets. Cofield and Fayetteville are located in the eastern part of North Carolina—a region with a significant concentration of swine operations. Candor and Norwood are located in central North Carolina, while Roaring River is in the western part of the state. The central and western regions have significant concentrations of poultry operations. The three corn markets and the Cofield soybean market are feed mills, while Fayetteville is a soybean processor and Norwood is a country elevator. Individual market volumes waxed and waned as the animal industry developed in North Carolina over the sample period, with some markets closing and others opening over time.[14] It should be noted that, in addition to barriers to trade and arbitrage among local markets, deviations from the law of one price in comparisons involving futures prices could also reflect low liquidity or other barriers to adjustment in futures markets.[15]

Figure 1 illustrates the time series of daily prices over the period of study. The prices are clearly closely related (often indistinguishable) and considerable periods of high volatility are noted in both series. We utilize nearby daily futures prices collected from the corn and soybean futures exchanges to measure the aggregate market. The futures prices are taken for the nearby (closest to expiration) contract and are rolled over to subsequent contracts on the last trading day of the month prior to expiration. Basis is defined as the local spot price minus the corresponding futures price. Basis tends to reflect local supply and demand conditions and thus may be low (or negative) immediately following harvest but may rise in the months prior to harvest, reflecting the relative scarcity of corn and soybeans. Differences in basis reflect the transactions costs for trading grain across local markets and basis levels, which are usually positive for these markets, and which tend to reflect the costs of importing grain from U.S. hinterland markets to North Carolina. The basis may fluctuate in response to local barriers to trade, such as transportation regulation and local logistical constraints, including storage capacity, as well as localized shocks reflecting regional shortages or surpluses of the commodity. Differences in seasonality across markets may also result in changes in local basis.

The first step in our analysis is to consider the individual time-series properties of the price data. Logarithmic prices, including the futures prices, were found to be highly nonstationary in every case. However, as would be expected, basis was found to be stationary in every case.[16] We then considered linkages among pairs of logarithmic prices. We considered comparisons for logarithmic prices without basis adjustments and for local market basis. Table 1 presents the results of an evaluation of price linkages, through regression models of pairs of logarithmic prices. The regression models confirm a very strong relationship among individual pairs of prices, with intercept terms close to zero and price transmission parameters very close to one. Further, the regression models display very high R^2 values, confirming the strength of daily spatial price linkages. Engel-Granger/Augmented Dickey-Fuller tests of the stationarity of the residuals were also conducted and strongly support the existence of a stable long-run relationship among pairs of prices. When the analysis is conducted for the basis series, the transmission elasticities are much smaller, indicating imperfect pass-through of basis shocks. The basis transmission elasticities range from 0.5 to 0.6 and the regression R^2 values are much smaller.

We then consider multivariate cointegration tests for the sets of logarithmic corn and soybean prices (three local markets and the relevant futures price). The results are presented in Table 2. Complete integration of markets would suggest that every pair of prices should be cointegrated, or that there should be three unique cointegrating vectors for the sets of

[13] In the case of Norwood soybeans, we also used a regression on the nearby market of Lumberton, which has nearly identical prices. In no case were more than 5% of the daily prices missing for any of the markets considered here.

[14] Maps illustrating the location of the markets and the spatial distribution of feeding operations are presented in the online supplementary appendix.

[15] We are grateful to an anonymous referee for pointing this out.

[16] Unit root tests are not presented here but are available in the online supplementary appendix.

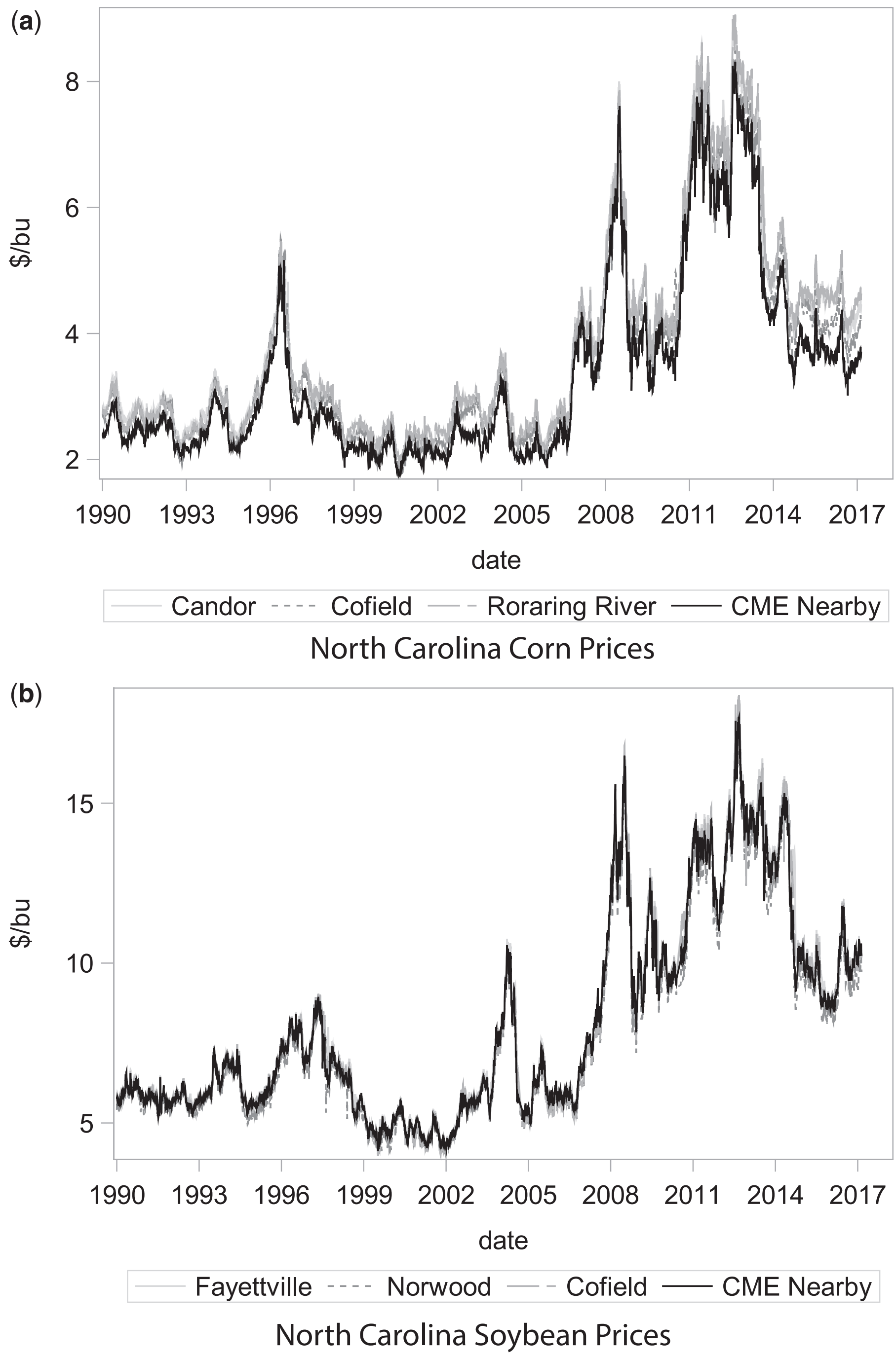


Figure 1. Corn and soybean prices

four prices. We considered both the trace and maximum Eigenvalue tests of Johansen (1991). In each case, the results indicate three unique cointegrating relationships and thus confirm that the local markets are linked together and that all of the markets are linked in the long-run to the futures prices.

The suite of nonlinearity tests described above was applied to the basis series. The testing results are presented in table 3. The tests strongly reject linearity in every case. The threshold models find two statistically-significant thresholds for each comparison. The multivariate TVAR models also reflect two

Table 1. Bivariate Transmission Elasticities and Cointegration Testing Results $P_t^1 = \alpha + \beta P_t^2 + \epsilon_t$

LHS Variable	RHS Variable	Parameter	Estimate	Standard Error	t Ratio	R Square	ADF Test[a]
Corn							
Candor	Cofield	α	0.0652	0.0016	40.62	0.9891	−9.05
		β	0.9873	0.0013	777.12		
Candor	RoaringRiver	α	0.0118	0.0015	7.96	0.9914	−10.63
		β	0.9895	0.0011	876.45		
Cofield	Roaring River	α	−0.0424	0.0021	−20.74	0.9838	−8.34
		β	0.9929	0.0016	637.30		
Candor	CME Nearby	α	0.1401	0.0017	83.84	0.9867	−8.52
		β	0.9865	0.0014	703.99		
Cofield	CME Nearby	α	0.0855	0.0020	42.14	0.9807	−7.49
		β	0.9907	0.0017	582.62		
Roaring River	CME Nearby	α	0.1375	0.0020	69.26	0.9815	−7.75
		β	0.9901	0.0017	594.61		
Candor-CME	Cofield-CME	α	0.0828	0.0007	121.39	0.4583	—
		β	0.5605	0.0075	75.12		
Candor-CME	R. River-CME	α	0.0458	0.0009	48.31	0.5483	—
		β	0.6258	0.0070	89.99		
Cofield-CME	R. River-CME	α	0.0011	0.0014	0.80	0.3282	—
		β	0.5848	0.0102	57.08		
Soybeans							
Fayetteville	Norwood	α	0.0284	0.0021	13.70	0.9934	−12.20
		β	1.0097	0.0010	983.38		
Fayetteville	Cofield	α	0.0411	0.0018	22.89	0.9950	−9.98
		β	0.9859	0.0009	1128.93		
Norwood	Cofield	α	0.0194	0.0018	10.65	0.9947	−11.39
		β	0.9730	0.0009	1100.66		
Fayetteville	CME Nearby	α	−0.0130	0.0022	−5.90	0.9929	−11.23
		β	1.0145	0.0011	946.91		
Norwood	CME Nearby	α	−0.0354	0.0020	−17.88	0.9940	−13.19
		β	1.0020	0.0010	1036.01		
Cofield	CME Nearby	α	−0.0504	0.0021	−24.06	0.9937	−10.31
		β	1.0269	0.0010	1004.29		
Fayetteville-CME	Norwood-CME	α	0.0337	0.0005	65.84	0.2432	—
		β	0.5539	0.0122	45.47		
Fayetteville-CME	Cofield-CME	α	0.0138	0.0003	45.78	0.4094	—
		β	0.6461	0.0097	66.78		
Norwood-CME	Cofield-CME	α	−0.0333	0.0003	−113.84	0.3020	—
		β	0.4941	0.0094	52.76		

Note: Superscript [a] indicates that basis (local price minus CME price) is stationary in every case. Thus, no cointegration tests are presented.

significant thresholds. In the case of the trivariate corn basis VAR model, the two thresholds were at {-0.0389, 0.0560}, which defines three regimes having 6.2%, 91.6%, and 2.2% of the observations in each regime. For soybean basis, we found optimal thresholds at {−0.0296, 0.0542} in the three-variable model, which defines three regimes of 2.2%, 95.8%, and 2.0% of the observations. These results indicate different patterns in adjustment, depending on the level of basis differences relative to the thresholds. These results are similar to those revealed in other threshold analyses, where large but infrequent deviations from parity trigger different patterns of adjustment and small differences correspond to limited adjustment and no spatial trade since transactions costs exceed the price (basis) differentials. The test results favor two thresholds over one, with one typically negative and another positive. This may reflect bidirectional trade between local markets, with commodity flowing from markets with lower prices to those with higher prices. In light of these findings, we estimate both a two-threshold VAR model and the semi-parametric VGAM models. Impulse response analysis is used to describe the patterns of adjustment implied by the estimated models.

Table 2. Multivariate Cointegration Testing Results[a]

H0: Rank = r	H1: Rank >r	Trace	Pr >Trace	Max Eigenvalue	Pr > Max
Corn					
0	0	290.59	0.0001	147.70	0.0001
1	1	142.89	0.0001	78.64	0.0001
2	2	64.24	0.0001	61.35	0.0001
3	3	2.89	0.0892	2.89	0.0892
Soybeans					
0	0	454.05	0.0001	208.59	0.0001
1	1	245.45	0.0001	139.85	0.0001
2	2	105.61	0.0001	102.70	0.0001
3	3	2.91	0.0881	2.91	0.0881

Note: Superscript [a] indicates that Johansen's test evaluates the rank of the cointegrating relationships among several variables. In the case of perfect integration of k markets, one would expect to find $k-1$ unique cointegating relationships. Our testing results confirm this and thus support spatially integrated markets.

The vector generalized additive vector autoregressive (VGAM) models were estimated using the penalized maximum likelihood procedures described above.[17] We assumed Gaussian distributions for the error structure. We also estimated a threshold vector autoregressive model using the same general specification and error structure. All lagged price effects are allowed to have fully nonparametric effects in representing price dynamics. Simple parametric intercept and dispersion terms are also included in the model. Consideration of Schwartz-Bayesian criteria indicated lag orders of 2–3. We choose a lag order of 2 for both models in light of the analytical complexity of the VGAM model, and to permit a straightforward comparison to the TVAR model. Tables 4 and 5 present the estimated parameters and implied degrees of freedom parameters for the VGAM models. Although the penalized likelihood function is quasi-maximum likelihood, a heuristic likelihood test of linearity can be derived by considering the difference in likelihood functions and degrees of freedom in a comparison to a linear VAR model. In every case (see tables 4 and 5), the associated likelihood ratio test statistic is highly significant.[18]

Tables 4 and 5 present the penalized VGAM estimates for corn and soybeans, respectively. The effective degrees of freedom (DF) represents the extent of nonlinearity implied by the responses. In the case of corn, significant nonlinearities are implied in degrees of freedom parameter estimates that approach the upper bound of nine and have corresponding small smoothing parameters. Most of the nonlinear spline effects are highly statistically significant. A linear model would imply 8 degrees of freedom in each equation. The effective degrees of freedom in each equation is significantly larger, typically in the 35–45 range.

The marginal effects of the lagged prices and thus the dynamic patterns of adjustment are best interpreted through a consideration of the smoothed response over the ranges of the explanatory factors. Figures 2 and 3 present the marginal effects and first derivatives that correspond to price transmission elasticities with 95% confidence intervals. In the case of linear responses, such as that in panel (b) of figure 2 (Candor's response to the lagged Cofield corn price), a constant price transmission elasticity is indicated. These responses are normalized about the mean values of the explanatory factors. The confidence bands are narrowest at the mean values and are, of course, wider for less statistically significant spline effects, as reflected in the F-statistics in tables 4 and 5. Note that the additive nature of the models does not allow explicit interaction terms among contemporaneously dated variables. However, the time-series structure of the VGAM models does allow for rich dynamics across the markets. As would be expected,

[17] We used the GAMPL procedure of SAS and the vecgam package in R, written by Yee (2017), to estimate the models. Threshold models were estimated using the R package tsDyn, written by Stigler (2018), and the MODEL procedure of SAS. Nonlinearity tests were conducted using the tseries package of R, written by Trapletti, Hornick, and LeBaron (2018).

[18] Note that the TVAR is not nested in the VGAM model and thus a likelihood ratio test cannot be derived. Parameter estimates for the TVAR models are not of particular interest in and of themselves and are not presented here but are included in the online supplementary appendix.

Table 3. Nonlinearity Specification Testing Results

	Corn					
	Candor/Cofield		Candor/Roaring River		Cofield/Roaring River	
Nonlinearity Test	Test Statistic	p-Value	Test Statistic	p-Value	Test Statistic	p-Value
LRT F-Test of AR vs. SETAR (1 Threshold)	381.21	0.0001	139.50	0.0001	94.29	0.0001
LRT F-Test of AR vs. SETAR (2 Thresholds)	392.98	0.0001	321.72	0.0001	88.83	0.0001
Teräsvirta's Neural Network χ^2	27.24	0.0001	194.59	0.0001	16.91	0.0002
White's Neural Network χ^2	45.13	0.0001	115.43	0.0001	29.82	0.0001
Tsay's Linearity Test F-Statistic	8.55	0.0001	5.80	0.0001	3.68	0.0001
3-Variable VAR System						
LRT F-Test of VAR vs. TVAR (1 Threshold)			687.93	0.0001		
LRT F-Test of VAR vs. TVAR (2 Thresholds)			878.01	0.0001		
	Soybeans					
	Fayetteville/Norwood		Fayetteville/Cofield		Norwood/Cofield	
Nonlinearity Test	Test Statistic	p-Value	Test Statistic	p-Value	Test Statistic	p-Value
LRT F-Test of AR vs. SETAR (1 Threshold)	63.49	0.0001	113.21	0.0001	160.61	0.0001
LRT F-Test of AR vs. SETAR (2 Thresholds)	136.03	0.0001	143.87	0.0001	186.04	0.0001
Teräsvirta's Neural Network χ^2	34.94	0.0001	28.47	0.0001	366.85	0.0001
White's Neural Network χ^2	23.74	0.0001	88.55	0.0001	132.67	0.0001
Tsay's Linearity Test F-Statistic	5.01	0.0001	5.07	0.0001	3.68	0.0001
3-Variable VAR System						
LRT F-Test of VAR vs. TVAR (1 Threshold)			232.87	0.0001		
LRT F-Test of VAR vs. TVAR (2 Thresholds)			353.14	0.0001		

Table 4. Penalized MLE GAM Model Estimates: Corn[a]

Component	Effective DF	Smoothing Parameter	Roughness Penalty	F Value
$Spline^1_{11}$	9.00000	<0.00001	<0.00001	2676.29*
$Spline^1_{21}$	1.11848	0.87890	<0.00001	1.70
$Spline^1_{31}$	8.99844	<0.00001	0.00003	411.51*
$Spline^1_{12}$	8.99559	<0.00001	0.00002	168.80*
$Spline^1_{22}$	1.85499	0.11640	0.00009	3.63*
$Spline^1_{32}$	5.16194	0.00059	0.00054	38.11*
α_1	1.00000	0.12481	0.00014	904.44*
σ_1	1.00000	0.00013	0.00018	7874.02*
Effective DF_1	37.12944			
Roughness Penalty_1	0.00069			
Likelihood Ratio Test (VAR)	1302*			
$Spline^2_{11}$	9.00000	<0.00001	<0.00001	273.81*
$Spline^2_{21}$	1.00154	10.08020	<0.00001	2938.99*
$Spline^2_{31}$	9.00000	<0.00001	<0.00001	380.63*
$Spline^2_{12}$	8.99644	<0.00001	<0.00001	54.10*
$Spline^2_{22}$	8.72627	<0.00001	0.00031	135.92*
$Spline^2_{32}$	8.99844	<0.00001	<0.00001	28.18*
α_2	1.00000	0.07496	0.00013	572.20*
σ_2	1.00000	0.00011	0.00016	0.71
Effective DF_2	47.72269			
Roughness Penalty_2	0.00031			
Likelihood Ratio Test (VAR)	908*			
$Spline^3_{11}$	8.99426	<0.00001	0.00009	321.39*
$Spline^3_{21}$	1.26905	0.45270	0.00003	0.19
$Spline^3_{31}$	9.00000	<0.00001	<0.00001	3642.15*
$Spline^3_{12}$	8.99419	<0.00001	<0.00001	72.26*
$Spline^3_{22}$	1.42236	0.28700	0.00006	0.11
$Spline^3_{32}$	8.84959	<0.00001	0.00036	208.16*
α_3	1.00000	0.12628	0.00015	870.92*
σ_3	1.00000	0.00014	0.00020	0.71
Effective DF_3	40.52944			
Roughness Penalty_3	0.00054			
Likelihood Ratio Test (VAR)	1054*			

Note: Superscript [a] indicates that the Spline ij_k terms represent the ith price effect, lagged j times, in the kth equation. The α_i and σ_i terms are the ith equation parametric intercept and dispersion terms and are presented with standard errors in place of the Roughness Penalty and t-ratios in place of F-Statistics. Asterisks indicate statistical significance at the $\alpha = .05$ or smaller level.

and in a manner consistent with standard linear VAR model parameters, the largest and most statistically significant effects occur in the once-lagged own values of each variable.

The derivatives indicate that significant nonlinearities appear to arise in the case of extreme basis changes.[19] This is consistent with the results of other nonparametric models of price responses, including those of Mancuso, Goodwin, and Grennes (2003) and Serra et al. (2006). In the case of corn markets, extreme lagged price decreases appear to trigger much greater corresponding

[19] Note that the derivatives are all presented on the same scale, whereas the marginal effects are equation-specific. This allows a consistent comparison of adjustments across markets and variables while still allowing the resolution to reveal nonlinear marginal effects.

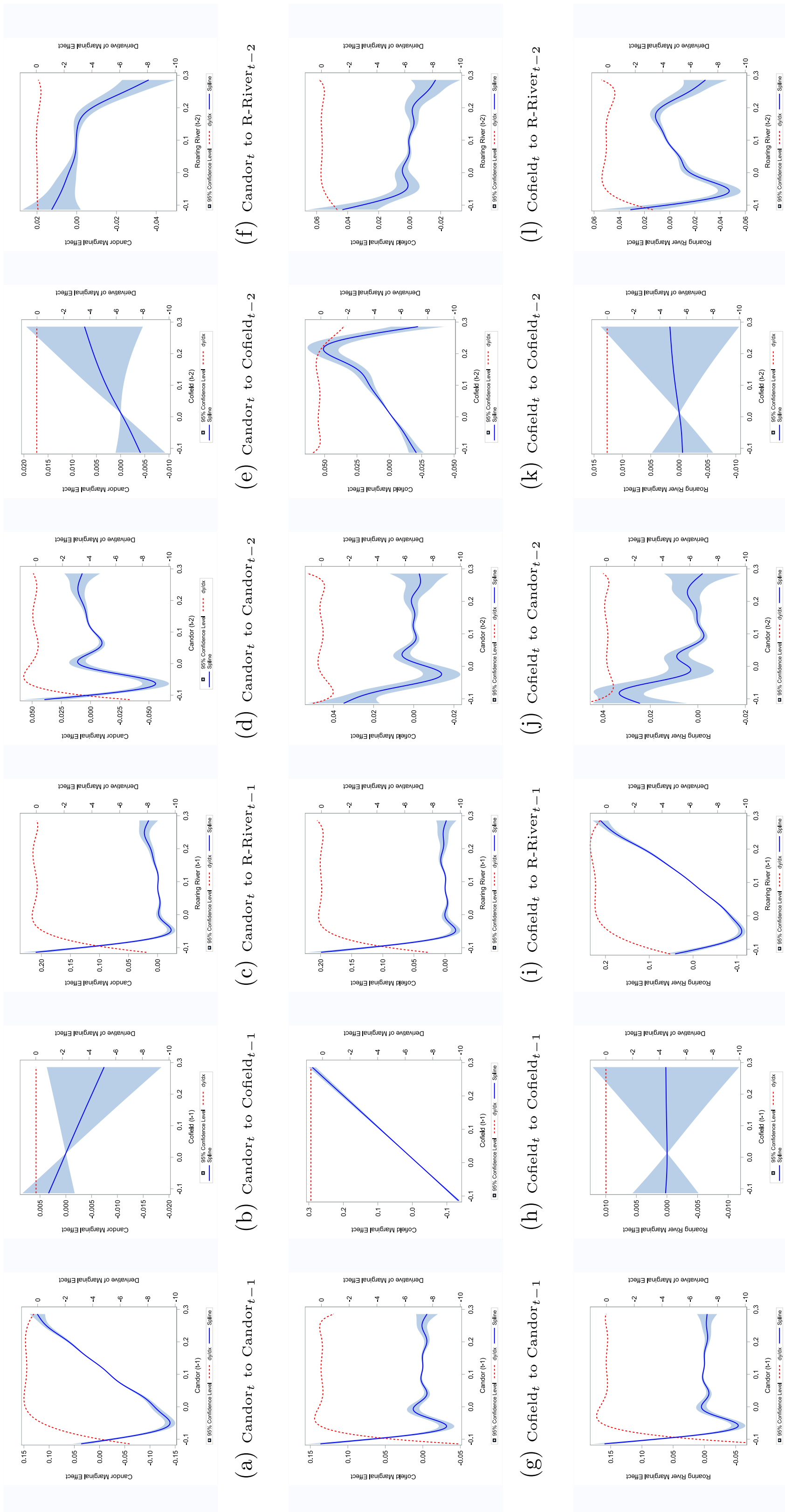
Table 5. Penalized MLE GAM Model Estimates: Soybeans[a]

Component	Effective DF	Smoothing Parameter	Roughness Penalty	F Value
$Spline^1_{11}$	9.00000	<0.00001	<0.00001	3116.27*
$Spline^1_{21}$	1.08940	0.14070	0.00006	0.82
$Spline^1_{31}$	1.00015	112.20000	<0.00001	0.71
$Spline^1_{12}$	7.66044	0.00003	0.00022	64.56*
$Spline^1_{22}$	9.00000	<0.00001	<0.00001	33.60*
$Spline^1_{32}$	2.85487	0.00554	0.00018	7.99*
α_1	1.00000	0.01630	0.00012	133.62*
σ_1	1.00000	0.00010	0.00014	10511.27*
Effective DF_1	32.60485			
Roughness Penalty_1	0.00046			
Likelihood Ratio Test (VAR) 134*				
$Spline^2_{11}$	4.95480	0.00043	0.00025	14.68*
$Spline^2_{21}$	8.93084	<0.00001	0.00024	2823.25*
$Spline^2_{31}$	1.01763	1.91510	0.00001	0.16
$Spline^2_{12}$	6.55792	0.00014	0.00040	28.69*
$Spline^2_{22}$	9.00000	<0.00001	<0.00001	189.83*
$Spline^2_{32}$	1.02826	1.19740	0.00002	1.29
α_2	1.00000	<0.00001	0.00014	−227.88*
σ_2	1.00000	0.00012	0.00017	0.71
Effective DF_2	33.48945			
Roughness Penalty_2	0.00091			
Likelihood Ratio Test (VAR) 400*				
$Spline^3_{11}$	1.00574	10.90290	<0.00001	38.74*
$Spline^3_{21}$	1.01478	2.91900	<0.00001	5.90*
$Spline^3_{31}$	8.99774	<0.00001	<0.00001	1902.72*
$Spline^3_{12}$	1.02309	2.70180	<0.00001	35.09*
$Spline^3_{22}$	1.01409	3.06160	<0.00001	7.69*
$Spline^3_{32}$	9.00000	<0.00001	<0.00001	179.96*
α_3	1.00000	0.00379	0.00012	32.68*
σ_3	1.00000	0.00009	0.00012	0.71
Effective DF_3	24.05543			
Roughness Penalty_3	0.00000			
Likelihood Ratio Test (VAR) 362*				

Note: Superscript [a] indicates that the Spline ij_k terms represent the ith price effect, lagged j times, in the kth equation. The α_i and σ_i terms are the ith equation parametric intercept and dispersion terms and are presented with standard errors in place of the Roughness Penalty and t-ratios in place of F-Statistics. Asterisks indicate statistical significance at the $\alpha = .05$ or smaller level.

responses in the current price than is the case for smaller changes and for changes corresponding to price increases. This is reflected in the substantial drop in the elasticities on the left portions of the response diagrams. Marginal responses for soybeans are similar but tend to indicate somewhat less nonlinearity in adjustments. In several cases, extremes in lagged basis tend to correspond to greater adjustments in current basis. This significant nonlinearity is typically apparent for both large and small basis changes (see, e.g., panels (h), (k), and (o) of figure 3). Marginal responses having confidence bands that are very wide and/or that fully encompass zero correspond to a lack of statistical significance. In most cases, own- and cross-price effects are statistically significant for values away from the mean.

The overall patterns of adjustment are best represented using dynamic, generalized impulse responses. Figure 4 presents dynamic



(a) Candor_t to Candor_{t-1} (b) Candor_t to Cofield_{t-1} (c) Candor_t to R-River_{t-1} (d) Candor_t to Candor_{t-2} (e) Candor_t to Cofield_{t-2} (f) Candor_t to R-River_{t-2}

(g) Cofield_t to Candor_{t-1} (h) Cofield_t to Cofield_{t-1} (i) Cofield_t to R-River_{t-1} (j) Cofield_t to Candor_{t-2} (k) Cofield_t to Cofield_{t-2} (l) Cofield_t to R-River_{t-2}

(m) R-River_t to Candor_{t-1} (n) R-River_t to Cofield_{t-1} (o) R-River_t to R-River_{t-1} (p) R-River_t to Candor_{t-2} (q) R-River_t to Cofield_{t-2} (r) R-River_t to R-River_{t-2}

Figure 2. Corn price VGAM model marginal effects and first derivatives

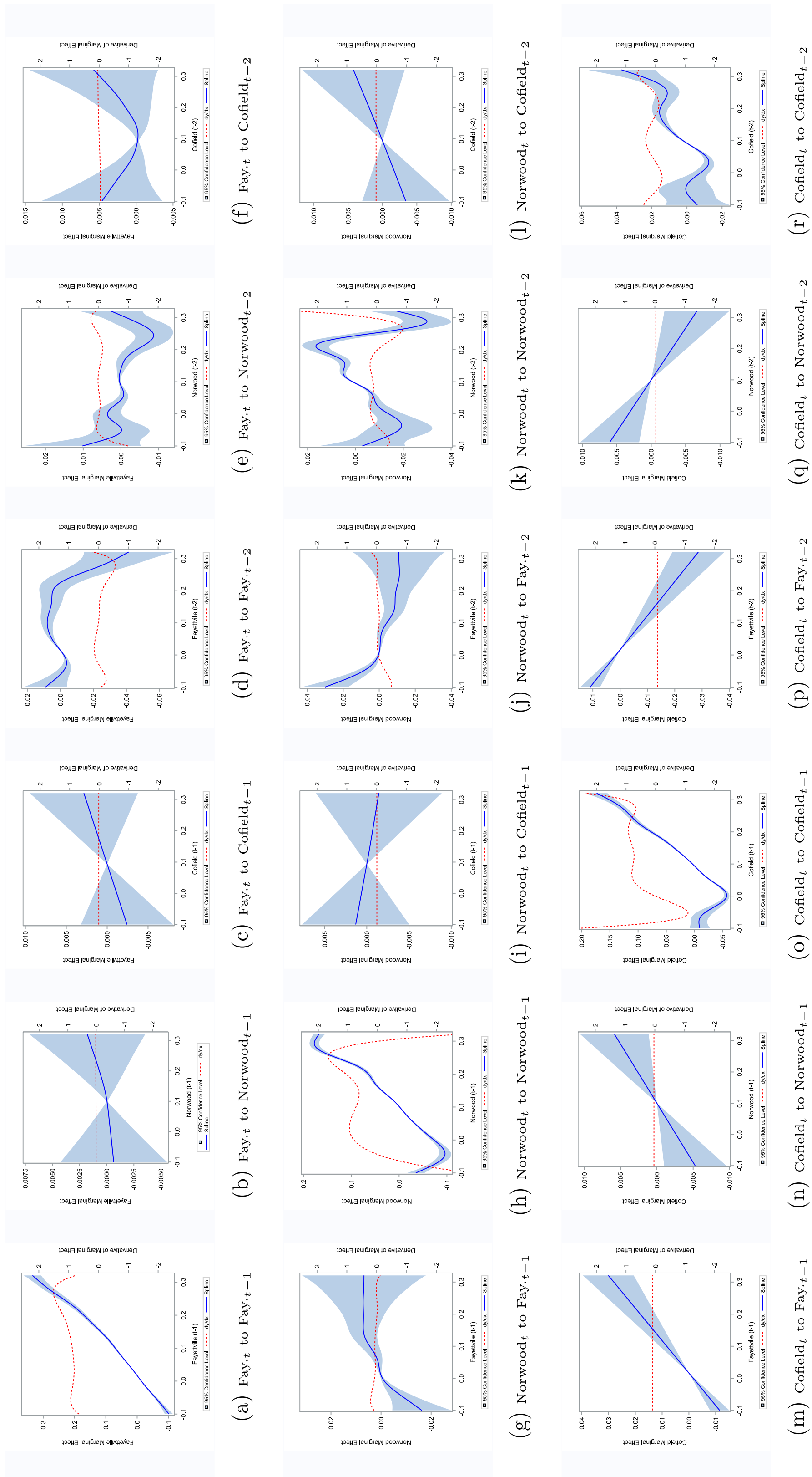
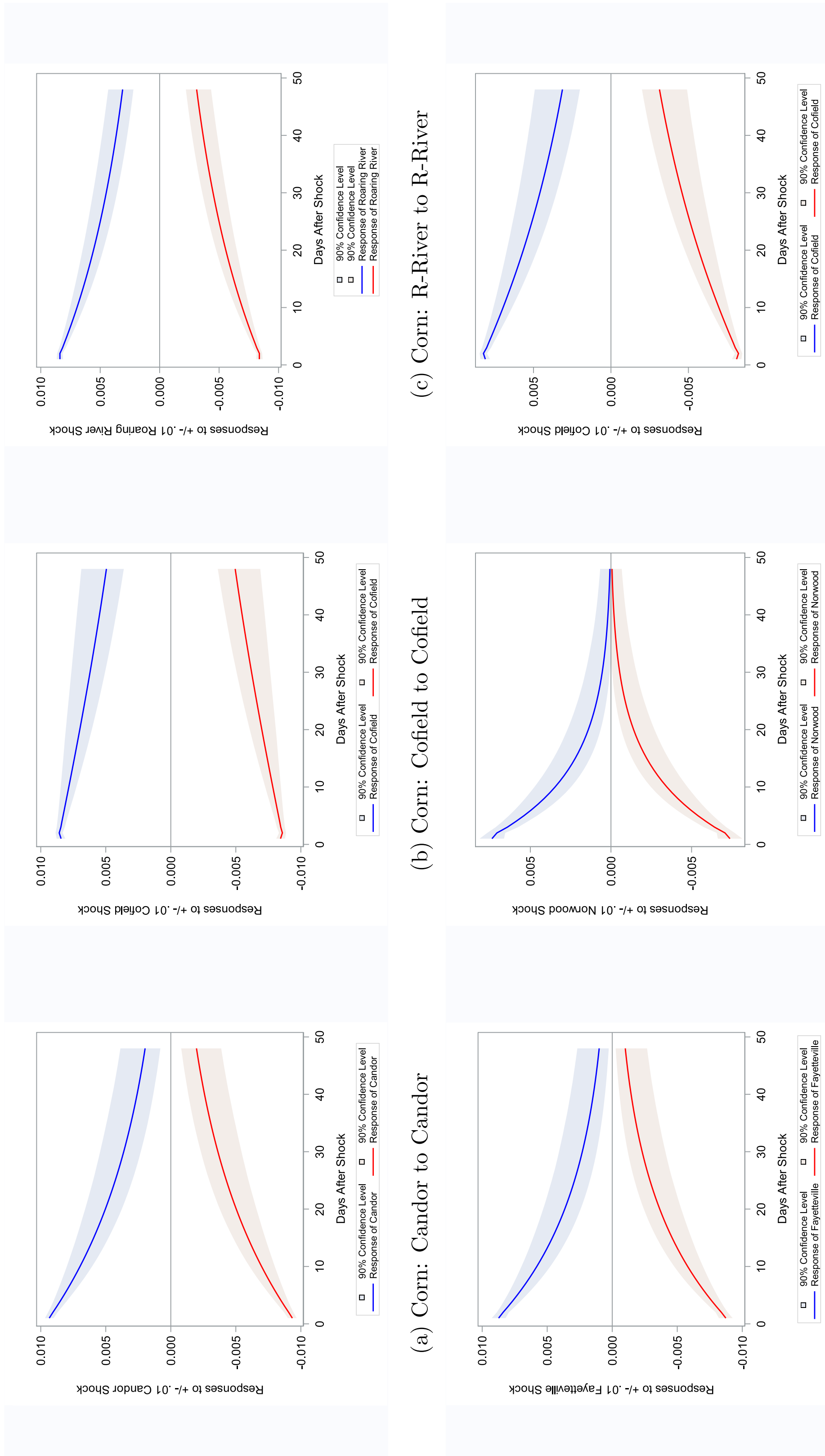


Figure 3. Soybean price VGAM model marginal effects and first derivatives



(a) Corn: Candor to Candor

(b) Corn: Cofield to Cofield

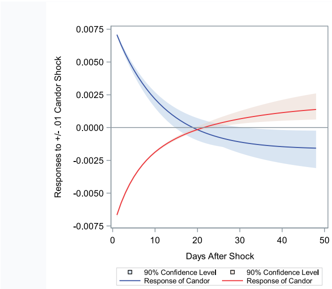
(c) Corn: R-River to R-River

(d) Soybeans: Fayetteville to Fayetteville

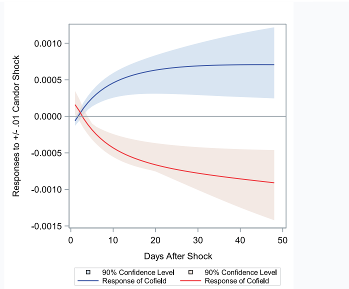
(e) Soybeans: Norwood to Norwood

(f) Soybeans: Cofield to Cofield

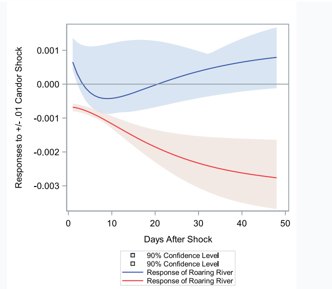
Figure 4. Threshold VAR model: own basis responses to 1% basis shocks



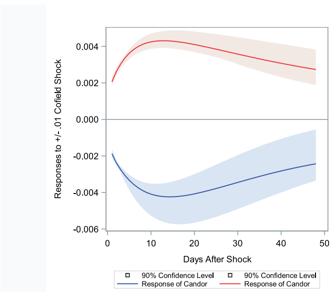
(a) Candor to Candor



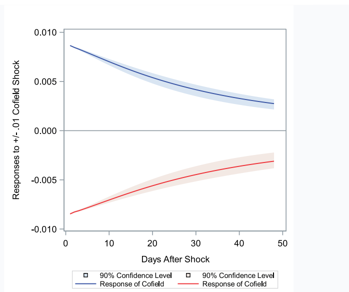
(b) Cofield to Candor



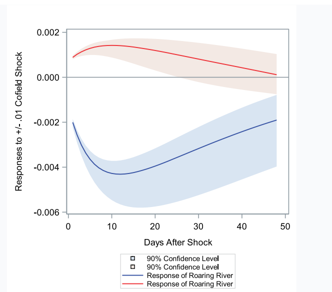
(c) R-River to Candor



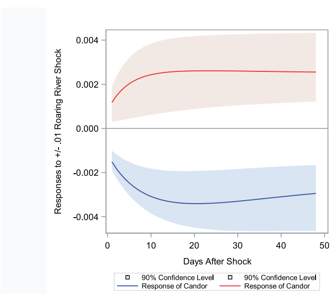
(d) Candor to Cofield



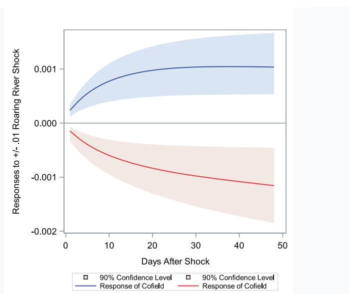
(e) Cofield to Cofield



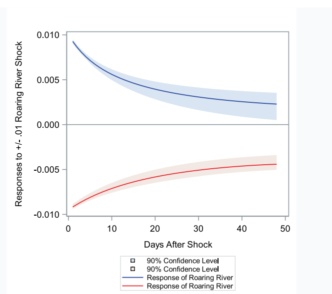
(f) R-River to Cofield



(g) Candor to R-River



(h) Cofield to R-River



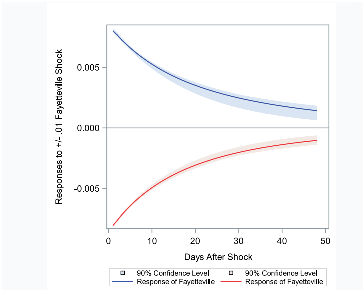
(i) R-River to R-River

Figure 5. VGAM model: corn price responses to 1% positive and negative corn price shocks

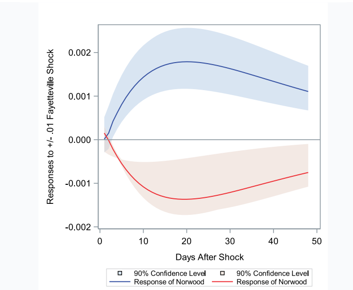
impulse responses for the TVAR models of corn and soybean basis.[20] As noted above, we consider dynamic impulse responses evaluated at the last observation for both the TVAR and VGAM model specifications. In the case of the TVAR model, the vast majority (greater than 90%) of the observations fall into the middle regime and thus the parameter estimates for this regime characterize the typical behavior of the markets' basis. This is not to imply that nonlinear responses may not be reflected in the impulse responses, but the responses to positive and negative 1% shocks appear symmetric, likely indicating that the responses apply to a single regime (i.e., that no regime change occurred over the 48-day period following the shock). The TVAR model responses are largely consistent with integration among the basis terms in that basis tends to settle to similar levels following shocks. However, the responses to basis shocks across markets (not presented here) are typically not statistically different from zero, which is not favorable toward market integration. Of course, as these results apply largely to the middle regime between very large and very small departures from a long-run equilibrium, one may not expect the responses to reveal significant adjustments.

Figures 5 and 6 present dynamic, generalized impulse responses calculated from the

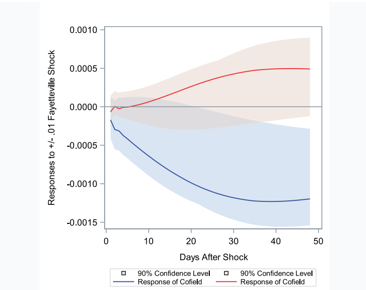
[20] Only own-basis responses are presented here. The entire set of impulse responses is presented in an online appendix.



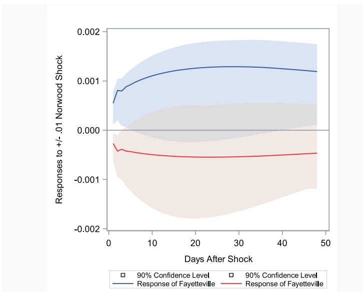
(a) Fayetteville to Fayetteville



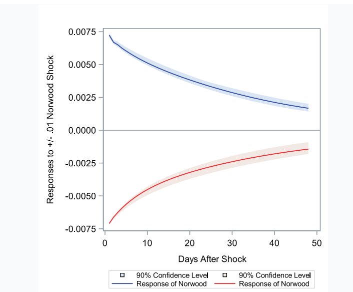
(b) Norwood to Fayetteville



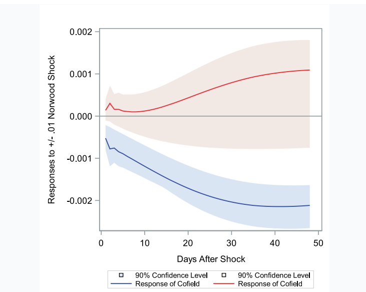
(c) Cofield to Fayetteville



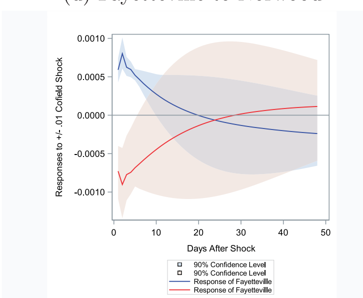
(d) Fayetteville to Norwood



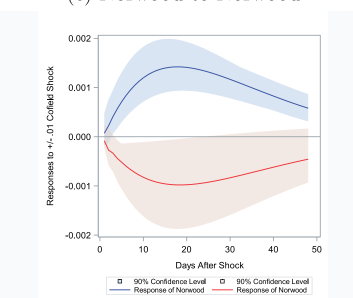
(e) Norwood to Norwood



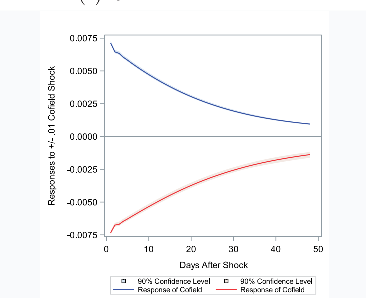
(f) Cofield to Norwood



(g) Fayetteville to Cofield



(h) Norwood to Cofield



(i) Cofield to Cofield

Figure 6. VGAM model: soybean price responses to 1% positive and negative soybean price shocks

VGAM semi-parametric models. Two results are immediately notable. First, substantial asymmetries in patterns of adjustment exist in response to 1% shocks of opposite sign. Even in those cases where the pattern of the response is symmetric, the extent of long-run adjustment is often quite different. For example, the Roaring River corn basis responses to Cofield shocks (panel (f) of figure 5) are of a similar shape but indicate a much greater response to a negative shock than to a positive shock of the same magnitude. Second, the results reveal a higher degree of statistical significance in the responses in that the confidence intervals are considerably more narrow and usually do not encompass zero. The results are largely consistent with integrated markets where shocks of a given size and magnitude to one market tend to evoke long-run equilibrating adjustments such that basis in other markets tends to settle at a similar level. The own-price responses are much more statistically significant than is the case for cross-price effects, a result that is consistent across all models considered. Of course, the greater degree of nonlinearity reflected in the semi-parametric models certainly suggests that shocks at different times and of different signs and magnitudes could reflect different patterns of adjustment.[21]

[21] Future research will certainly benefit from considering generalized impulse responses evaluated at all observations. The very long computational time associated with estimation and bootstrapping necessarily limits the impulse response analysis.

In sum, our results are largely consistent with prior research that has examined many of these same markets, albeit over different periods and without the adjustments for futures market prices. Consistent with the results of Goodwin and Piggott (2001), Sephton (2003), and Greb et al. (2013), we find evidence of significant nonlinearities in patterns of spatial basis adjustment in local markets for corn and soybeans in North Carolina. Impulse responses indicate that the responses are largely symmetric in the case of threshold VAR models. Conversely, varying degrees of nonlinearity and asymmetry are indicated by the VGAM models. We view the application of generalized additive vector autoregressive models to be a natural next step in the growing body of literature that addresses nonlinear patterns of price adjustments and market integration.

Concluding Remarks

Following a long line of research into nonlinear time-series models of spatial price parity and market integration, we apply semi-parametric, generalized additive vector autoregressive models. We also consider threshold models, which have been frequently applied to analysis of spatial market integration that are compared to the fully flexible, vector generalized additive autoregressive models. Our application is to basis series for three North Carolina corn and three North Carolina soybean markets. We find significant evidence of nonlinear patterns of adjustment across markets in response to exogenous shocks to basis. Marginal effects in responses and corresponding derivatives (price/basis transmission elasticities) are typically nonlinear and suggest greater adjustment to extreme basis shocks. Generalized impulse response functions provide evidence in favor of efficiently linked markets and nonlinear responses. In particular, exogenous shocks tend to lead to relatively homogeneous adjustments to basis in spatially-separated markets. The semi-parametric VGAM models generally reveal more statistical significance and substantially more nonlinearity in adjustments than is the case for the threshold VAR models.

Future research may benefit from considering the impacts of specific shocks at specific periods of time in the data, or from considering an aggregate of responses across all observations. A number of nonparametric spline alternatives could be used to define richer (more flexible) empirical models. In particular, multivariate splines allowing for contemporaneous interactions and thin plate splines that are not dependent upon knots could be used to potentially expand the scope of the VGAM models. Likewise, implications of market basis integration for spatial linkages also raises a number of questions, including seasonality, variable transportation costs, and structural breaks. These issues remain important topics for future research.

Supplementary Material

Supplementary material are available at *American Journal of Agricultural Economics* online.

References

Bai, J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66 (1): 47–78.

———. 2003. Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics* 18 (1): 1–22.

Balke, N.S., and T.B. Fomby. 1997. Threshold Cointegration. *International Economic Review* 38(3): 627–45.

Bates, D., M. Lindstrom, G. Wahba, and B. Yandell. 1987. GCVPACK—Routines for Generalized Cross Validation. *Communications in Statistics - Simulation and Computation* 16: 263–97.

Becker, R., W. Enders, and S. Hurn. 2004. A General Test for Time Dependence in Parameters. *Journal of Applied Econometrics* 19 (7): 899–906.

Becker, R., W. Enders, and J. Lee 2006. A Stationary Test with an Unknown Number of Smooth Breaks. *Journal of Time Series Analysis* 27: 381–409.

Enders, W., and M.T. Holt. 2012. Sharp Breaks or Smooth Shifts? An Investigation of the Evolution of Primary Commodity Prices. *American Journal of Agricultural Economics* 94 (3): 659–73.

———. 2014. The Evolving Relationships between Agricultural and Energy Commodity Prices A Shifting-Mean

Vector Autoregressive Analysis. *Chapter in The Economics of Food Price Volatility,* eds. Jean-Paul Chavas, David Hummels, and Brian D. Wright, 135–87. Chicago: University of Chicago Press.

Fackler, P., and B. Goodwin. 2001. Spatial Price Analysis. In *Handbook of Agricultural Economics*, ed. B.L. Gardner and G.C. Rausser, 972–1025. Amsterdam: Elsevier.

Gonzalez, A., and T. Teräsvirta. 2008. Modelling Autoregressive Processes with a Shifting Mean. *Studies in Nonlinear Dynamics and Econometrics* 12 (1) doi: 10.2202/1558-3708.1459.

Goodwin, B.K.B., and N.E. Piggott. 2001. Spatial Market Integration in the Presence of Threshold Effects. *American Journal of Agricultural Economics* 83 (2): 302–17.

Goodwin, B.K., M.T. Holt, and J. Prestemon. 2011. North American Oriented Strand Board Markets, Arbitrage Activity, and Market Price Dynamics: A Smooth Transition Approach. *American Journal of Agricultural Economics* 93 (4): 993–1014.

Gospodinov, N., A.M. Herrera, and E. Pesavento. 2013. Unit Roots, Cointegration, and Pretesting in Var Models. In *Advances in Econometrics, Vol.* 32, VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims, ed. T.B. Fomby, R.C. Hill, I. Jeliazkov, J.C. Escanciano, and E. Hillebrand. Bingley, UK: Emerald Group Publishing Limited.

Greb, F., S. von Cramon-Taubadel, T. Krivobokova, and A. Munk. 2013. The Estimation of Threshold Models in Price Transmission Analysis. *American Journal of Agricultural Economics* 95 (4): 900–16.

Hastie, T., and R. Tibshirani. 1986. Generalized Additive Models. *Statistical Science* 1 (3): 297–318.

Holt, M.T., and T. Teräsvirta. 2012. Global Hemispheric Temperature Trends and Co-trending: A Shifting Mean Autoregressive Analysis. CREATES Research Paper, 54. Department of Economics and Business Aarhus University.

Hornik, K., M. Stinchcombe, and H. White. 1989. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2 (5): 359–66.

Johansen, S. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59 (6): 1551–80.

Kilian, L., and H. Lütkepohl 2017. *Structural Vector Autoregressive Analysis.* Cambridge UK: Cambridge University Press.

Koop, G., M. Pesaran, and S. Potter. 1996. Impulse Response Analysis in Nonlinear Multivariate Models. *Journal of Econometrics* 74: 119–47.

Linton, O. 2000. Efficient Estimation of Generalized Additive Nonparametric Regression Models. *Econometric Theory* 16 (4): 502–23.

Lo, M.C., and E. Zivot. 2001. Threshold Cointegration and Nonlinear Adjustment to the Law of One Price. *Macroeconomic Dynamics* 5 (4): 533–76.

Mancuso, A.J., B.K. Goodwin, and T.J. Grennes. 2003. Nonlinear Aspects of Capital Market Integration and Real Interest Rate Equalization. *International Review of Economics and Finance* 12: 283–303.

Michael, P., A. Nobay, and D. Peel. 1994. Purchasing Power Parity Yet Again: Evidence from Spatially Separated Commodity Markets. *Journal of International Money and Finance* 13 (6): 637–57.

Ng, S., and T.J. Voselgang. 2002. Analysis of Vector Autoregressions in the Presence of Shifts in Mean. *Econometric Reviews* 21 (3): 353–81.

Nychka, D. 1988. Bayesian Confidence Intervals for Smoothing Splines. *Journal of the American Statistical Association* 83: 1134–43.

O'Connell, P., and S. Wei. 2002. The Bigger They Are, the Harder They Fall: Retail Price Differences across U.S. Cities. *Journal of International Finance* 56: 21–53.

Obstfeld, M., and A. Taylor. 1997. Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited. *Journal of the Japanese and International Economies* 11 (4): 441–79.

Park, H., J. Mjelde, and D. Bessler. 2007. Time–Varying Threshold Cointegration and the Law of One Price. *Applied Economics* 39 (9): 1091–105.

Piggott, N., G. Bullen, J. Dunphy, W. Everman, and D. Washburn. 2017.

Soybean Production and Marketing in North Carolina. NC Extension Publication AG-835. Available at: https://content.ces.ncsu.edu/north-carolina-soybean-production-guide/soybean-production-and-marketing-in-north-carolina. Accessed April 1, 2018

Sephton, P. 2003. Spatial Market Arbitrage and Threshold Cointegration. *American Journal of Agricultural Economics* 85 (4): 1041–6.

Serra, T., K. Goodwin, J. Gil, and A. Mancuso. 2006. Non-parametric Modelling of Spatial Price Relationships. *Journal of Agricultural Economics* 57 (3): 501–22.

Stigler, M. 2018. tsDyn: Nonlinear Time Series Models with Regime Switching. R Package version 0.9-46.

Stone, C. 1985. Additive Regression and Other Nonparametric Models. *Annals of Statistics* 13 (2): 435–44.

Taylor, A. 2001. Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean–Reversion Tests of the Law of One Price. *Econometrica* 69 (2): 473–98.

Teräsvirta, T, C. Lin, and C. Granger. 1993. Power of the Neural Network Linearity Test. *Journal of Time Series Analysis* 14: 209–20.

Trapletti, A., K. Hornik, and B. LeBaron. 2018. tseries: Time Series Analysis and Computational Finance: R Package Version 0.10-44.

Tsay, R. 1989. Testing and Modeling Threshold Autoregressive Processes. *Journal of the American Statistical Association* 84: 231–40.

Wahba, G. 1990. Spline Models for Observational Data. *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.

White, H. 1989. An Additional Hidden Unit Tests for Neglected Nonlinearity in Multilayer Feedforward Networks. In *Proceedings of the International Joint Conference on Neural Networks, Washington DC*, 451–55. New York: IEEE Press, 2.

Wood, S. 2003. Thin Plate Regression Splines. *Journal of the Royal Statistical Society, Series B* 65: 95–114.

Wood, S.N. 2004. Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistics Association* 99: 673–86.

Yee, T. 2015. *Vector Generalized Linear and Additive Models With an Implementation in R.* New York: Springer.

———. 2017. VGAM: Vector Generalized Linear and Additive Models. package version 1.0–3.

Yee, T., T. Hastie. 2003. Reduced–Rank Vector Generalized Linear Models. *Statistical Modelling* 3: 15–41.

Yee, T., C. Wild. 1996. Vector Generalized Additive Models. *Journal of the Royal Statistical Society (B)* 58: 481–93.