A gravity model approach to forecasting tuberculosis transmission in cattle

Fang Xie, Richard D. Horan, and Christopher A. Wolf

Department of Agricultural, Food and Resource Economics
Michigan State University

Selected Paper prepared for presentation at the American Agricultural Economics

Association Annual Meeting, Milwaukee, Wisconsin, July 2009

The authors gratefully acknowledge funding provided by the Economic Research Service-USDA cooperative agreement number 58-7000-6-0084 through ERS' Program of Research on the Economics of Invasive Species Management (PREISM), and the National Research Initiative Competitive Grants Program, USDA, CSREES, grant #2006-55204-17459. The views expressed here are the authors' and should not be attributed to ERS or USDA.

Abstract

Bovine tuberculosis (bTB) in cattle has caused significant economic losses to livestock producers and has proven difficult to eradicate. It is suspected that cattle movement across different farms and regions is one of the key factors of bTB transmission in the United States. Prior attempts to model the epidemiology of bTB infection within cattle to predict disease transmission have not adequately captured the behavioral aspects of trade. A better understanding of livestock trade patterns would help in predicting disease transmission and the associated economic effects. In this paper, we develop a gravity model of livestock trade and link it to an epidemiological model of bTB transmission, with the goal being that this information could lead to improved disease surveillance and management. Our findings suggest that feedbacks between jointly determined disease dynamics and trade system matter and should be considered together for efficient disease management.

Introduction

Tuberculosis (TB) is a common and often deadly infectious mycobacterium disease that occurs in both animals and humans. Bovine TB (bTB) in cattle has caused significant economic losses to livestock producers (with total indemnity costs exceeding $29 million in FY 2007) and has proven difficult to eradicate. It is suspected that cattle movement across different farms and regions is one of the key factors of bTB transmission in the United States. A better understanding of livestock trade patterns would help in predicting disease transmission and the associated economic effects. The purpose of this paper is to link a gravity model of livestock trade with an epidemiological model of bTB transmission, with the goal being that this information could lead to improved disease surveillance and management.

The gravity model has become the workhorse model to analyze patterns of trade in international economics. Gravity models were originally inspired by Newton's Law of Gravitation in physics, which suggests that gravity depends positively on mass and negatively on distance. The basic idea is that larger places (in terms of population or economic size) attract people, ideas, and commodities more than smaller places, and places closer together have a greater attraction. Gravity models represent one of the most empirically successful models in economics (Anderson and van Wincoop 2003), yielding sensible parameter estimates and explaining a large part of the variation in bilateral trade (Rose 2004). The theoretical foundation for these models has been developed by Anderson (1979), Anderson and van Wincoop (2003), and Eaton and Kortum (2002).

Ecologists have recently applied non-behavioral forms of gravity models to invasive species problems to estimate long-distance dispersal of species between discrete

points in heterogeneous landscapes, a problem similar (at least in some respects) to disease transmission. Bossenbroek et al. (2001, 2007) developed a production-constrained gravity model to forecast zebra mussel dispersal into inland lakes as a function of site characteristics, the relative locations of lakes, and the number and location of boats on which zebra mussels may hitch a ride. While their analyses show promise, there is no behavioral model of boat movements. Rather, the estimates are based on lake characteristics and do not consider the explicit economic incentives of boat owners to travel from one lake to another.

Prior attempts to model the epidemiology of bTB infection within cattle to predict disease transmission have not adequately captured the behavioral aspects of trade. For instance, Barlow et al. (1997) apply both deterministic and stochastic models to investigate bTB dynamics within cattle herds in New Zealand (Barlow et al. 1997). The simulations suggested that bTB was unlikely to persist in a herd, under present bTB testing policies, without an external source of reinfection. The most likely source of infection came from the movement of infected cattle into uninfected herds, and from surrounding wildlife species (Barlow et al. 1998). Cattle movements in these models are considered to occur taking human behaviors as exogenous and fixed. But trade is a behavioral phenomenon and so behavior plays a key role in cattle movement and hence disease transmission. Each year, tens of millions of cattle are shipped into anther state for feeding or breeding. Therefore, trade mechanisms must be integrated into epidemiology models to make transmission risks endogenous. That is, both epidemiology and economics are important and jointly determine bTB transmission patterns.

In this paper, we develop a gravity model to capture the economic incentives for cattle movement in US, and we tie this to an epidemiology model to predict disease risks. In the gravity model, the cattle movement and disease risks are driven by production costs, transportation costs that are increasing in the distances between buyers and sellers, and costs caused by trade restrictions imposed on regions that have lost TB-free disease status. The model is currently set up at the state level. Following Barlow et al. (1998), cattle herds in each state are divided into one of the three types: herds that are infected with bTB and identified as such, herds that are infected but not detected, and herds that are healthy but susceptible. Trade is between herds that are either susceptible or non-detected, and this cattle movement feeds into the epidemiology model to affect disease transmission. Trade restrictions may be introduced as new infected herds become identified, representing an epidemiological feedback that influences trade patterns in the gravity model.

A gravity model of trade

Assume cattle (a homogenous good) are produced in each state. Cattle are generally bred and partially raised in one location and later moved to a final location for fattening before slaughter. Cattle trade in our model refers to this movement. Our unit of analysis for the gravity model is a US state, though there are many buyers and sellers in each state and movement can occur within or across states. We index buyers by *j* and sellers by *i*, with the index referring to the state in which these producers reside. Sellers in state *i* produce cattle with an average input cost of c_i. Within state *i*, there is heterogeneity in the efficiency of production. Denote this efficiency by z_i, so that effective costs are c_i / z_i.

The efficiency parameter is taken to be random and is assumed to follow a Frechet distribution function (a special case of the generalized extreme value distribution, also called the Type II extreme value distribution), following Eaton and Kortum (2002):[1]

(1) $F_i(z) = e^{-z^{-\theta}}$

The distribution is assumed to be independent across states. The parameter $\theta > 0$ is a heterogeneity coefficient, with a bigger θ implying lower productivity differences across states. Specifically, z_i has geometric mean $e^{\tau/\theta}$, where τ is the Euler constant (= 0.577…), and $\ln(z_i)$ has a standard deviation $\pi/(\theta/6)$, where π is the constant pi (= 3.14…).

The term c_i / z_i represents the unit cost of production only, and does not reflect transportation and other trade related costs that would be relevant to the buyer's purchase decision. We focus on two kinds of trade costs: (i) transportation costs, and (ii) additional costs imposed by trade restrictions due to bTB. Specifically, the trade cost associated with moving cattle from state *i* to state *j* is denoted b_{ji}, where b_{ji} is a function that takes the following form

(2) $b_{ji}(d_{ji}, \delta, \omega) = (1+\delta)^{\gamma} d_{ji}^{\rho\omega}$

Here, δ is a dummy variable for TB-free disease status: δ = 0 if the selling state is TB-free, and δ = 1 if the state has lost its TB-free status. The exponent γ > 0 is a parameter. Hence, trade costs are higher when there is a trade restriction than when there is no restriction. The variable d_{ji} is the distance between two states. The term ω is a dummy

[1] The Frechet distribution is one of the common distribution models that can be applied when we generate N data sets from the same distribution, and create a new data set that includes the maximum values from these N data sets. Eaton and Kortum (2002) assume $F_i(z) = e^{-Y_i z^{-\theta}}$, with a higher Y_i meaning a higher average realization for region *i*. In contrast, we treat Y_i the same across states since the technology of cattle production does not vary substantially among cattle producers in the United States.

variable: ω = 0 if trade occurs within the same state, and ω = 1 if trade occurs between different states. Finally, ρ > 0 is a parameter. Under this specification, transportation costs are higher if the two regions trading are farther away.

After taking trade costs into account, the price for buyers in state *j* to buy one cattle produced in state *i* is the unit production cost multiplied by trade costs

(3) $$p_{ji} = c_i b_{ji} / z_i$$

Under perfect competition, p_{ji} would be the price buyer in state *j* would pay if it chose to buy cattle from state *i*. Buyers in state *j* would try to pay the lowest price across all sources, so the price they would pay would be:

(4) $$p_j = \min\{p_{ji}, i = 1,...N\}$$

where *N* is the total number of states.

Substituting the expression for p_{ji} into (4) results in the following distribution for price p_{ji}:

(5) $$G_{ji}(p) = PR[p_{ji} \le p] = 1 - F_i(c_i b_{ji} / p) = 1 - e^{-(c_i b_{ji})^{-\theta} p^{\theta}}$$

The distribution of prices at which state *j* will buy is

(6) $$G_j(p) = 1 - \prod_{i=1}^{N} [1 - G_{ji}(p)] = 1 - e^{-\Phi_j p^{\theta}}$$

where $\Phi_j = \sum_{i=1}^{N} (c_i b_{ji})^{-\theta}$.

The probability that a producer in state *i* sells a cattle to at the lowest price to a buyer in state *j* is simply:

(7) $$\phi_{ji} = PR[p_{ji} \le \min\{p_{js}, s \ne i\}] = \int \prod_{s \ne i} (1 - G_{js}(p) dG_{ji}(p)) = \frac{(c_i b_{ji})^{-\theta}}{\Phi_j}$$

where ϕ_{ji} is the fraction of cattle that producers in state *j* buy from producers in state *i*:

$$(8) \qquad \phi_{ji} = \frac{x_{ji}}{x_j} = \frac{(c_i b_{ji})^{-\theta}}{\Phi_j} = \frac{(c_i b_{ji})^{-\theta}}{\sum_{k=1}^{N} (c_k b_{jk})^{-\theta}},$$

where x_j is state *j*'s total cattle purchase and x_{ji} is the number of cattle that state *j* buys from state *i*.

We can express x_{ni} in the following form:

$$(9) \qquad x_{ji} = \frac{x_j \left(c_i b_{ji}(d_{ji}, \gamma)\right)^{-\theta}}{\sum_{k=1}^{N} \left(c_k b_{jk}(d_{jk}, \gamma)\right)^{-\theta}}$$

Equation (9) is our gravity equation. It tells us that the cattle movement, after controlling for size (total cattle purchases), depends negatively on both state *i*'s input costs and the trade costs between state *i* and state *j*, relative to the sum of an non-linear function of both input and trade costs across all states.

Equation (9) looks very similar to the standard gravity model in the ecological literature (Bossenbroek et al. 2001, 2007):

$$(10) \qquad U_{ji} = \frac{H_i W_j d_{ji}^{-\rho}}{\sum_{k=1}^{N} W_j d_{jk}^{-\rho}}$$

where U_{ij} is the flow of human activity movements from region *i* to region *j*, H_i is the number of people at location *i*, and W_j is the attractiveness of location *j*. However, there is no behavioral model for the human activity movements. The attractiveness of a location only depends on a couple of site characteristics which are invariant over time, rather than people's explicit economic incentives of those activities.

Estimation of the gravity model

Let $P_j = [\sum_{k=1}^{N} \left(c_k b_{jk}\right)^{-\theta}]^{-1/\theta}$, which could be regarded as a "multilateral resistance" index, as it depend on all the trade cost variables { b_{ji} }, including those not directly involving *j*. Then equation (9) could be rewritten as

(11) $$x_{ji} = x_j \left(\frac{c_i b_{ji}}{P_j} \right)^{-\theta}$$

We can estimate equation (11) by taking the log of both sides:

(12) $$\ln x_{ji} = \ln x_j + \theta \ln P_j - \theta \ln c_i - \theta \ln b_{ji}$$

Equation (12) is still nonlinear in θ since P_j is a nonlinear function of θ. Anderson and van Wincoop (2004) suggest a technique that provides consistent estimates of (12), and it has been adopted by Anderson and van Wincoop (2003), Eaton and Kortum (2002), and Rose and van Wincoop (2001). Consider a dummy variable, which Anderson and van Wincoop (2004) refer to as an outward region specific dummy, that indicates whether a state is a net importer or a net exporter of cattle. Specifically, let O_j be the outward region specific dummy: O_j = 1 if state *j*'s sales are greater than its purchases, and O_j = 0 if otherwise. Anderson and van Wincoop (2004) show that replacing $\ln x_j + \theta \ln P_j$ with the outward region specific dummy will give us consistent estimates using ordinary least squares. The gravity equation that we will estimate then is

(13) $$\ln x_{ji} = O_j - \theta \ln c_i - \theta\rho \ln(d_{ji}^{\omega}) - \theta\gamma \ln(1+\delta),$$

Data for the cattle movement variable x_{ji} comes from interstate livestock movement data from the USDA Economic Research Service (Shields and Mathews, Jr. 2003). There are cattle movement flows among 48 US states (Hawii and Alaska are not

included in this analysis). The outward region specific dummy O_j is coded by calculating the difference of cattle sales and purchases for state *j*. The input costs, c_i, are feed prices in dollars per hundredweight, as these represent the most important expense of raising cattle. The input costs were constructed using average corn and hay prices based on the feed proportion of a typical cattle farm (59% corn and 41% hay). The ration-based cost index is: [price of corn($/bu)/56 × 0.59 + price of hay($/ton)/2000 × 0.41] × 100. Distance between state *i* to state *j* (d_{ji}) is measured in kilometers as the distance between the center points of the two states. Finally, the dummy variable δ is coded based on a state's TB status. Currently, only three states have lost TB-free status—Michigan, Minnesota, and New Mexico, so δ =1 for these three states, and 0 otherwise.

One complication arising from the log-linearized version of the gravity model is that there are many zero trade flow observations in the data set and ln(0) is undefined. We must either address this issue or else drop all observations of zero trade flows from the sample. Disregarding zeros means we are getting rid of potentially useful information and we might be producing biased estimates of the coefficients we are primarily interested in.

There are several approaches have been applied in the literature to solve this problem. One is the so called "Ad Hoc" approach. Although ln(0) is not defined, ln(0+ε) is defined and can be used to approximate ln(0) for a very small value of ε. Therefore, adding a small and positive number to all trade flows can be a sensible place to start, to see if including or excluding zeros appears to make much of a difference in the estimation. This "Ad Hoc" approach has been commonly applied in the policy literature (Eichengreen and Irwin 1998), but it has no theoretical basis.

Another approach that has been applied extensively in gravity models in the international trade literature is Heckman's sample selection model (Heckman 1979, Emlinger 2008). This is the approach we adopt.

First, a set of covariates is used to determine the probability that two states engage in trade (i.e., that they are in the sample). The selection mechanism is

(14) $y_{ji} = O_j - \theta \ln c_i - \theta\rho \ln d_{ji}^{\omega} - \theta\gamma \ln(1+\delta) + x_j + \mu_{ji}$

$s_{ji} = 1$ if $y_{ji} > 0$

$s_{ji} = 0$ if $y_{ji} \leq 0$,

where s_{ji} is a selection dummy. The selection equation (14) determines whether or not we observe cattle trade between two sates in the sample, i.e, $\ln x_{ji}$ exists or can be dropped from subsequent estimations.

Next, a second set of covariates determines the intensity of bilateral trade, subject to the existence of a trade relationship. The regression model for this relation is specified as:

(15) $\ln \tilde{x}_{ji} = O_j - \theta \ln c_i - \theta\rho \ln d_{ji}^{\omega} - \theta\gamma \ln(1+\delta) + \varepsilon_{ji}$,

where $\ln x_{ji} = \ln \tilde{x}_{ji}$ if $s_{ji} = 1$, and $\ln x_{ji}$ = not observed (i.e., the observation is not used) if $s_{ji} = 0$. The error terms for the two equations are $(\mu_{ji}, \varepsilon_{ji})$~bivaraite normal [0, 0, 1, $\sigma_{\varepsilon}^2, \zeta_{\varepsilon\mu}$].

We first run a Heckman two-step model (equation (14) and (15)) to test whether there is sample selection. The p_value for the inverse mills ratio is 0.055, so we cannot reject the null hypothesis that there is no sample selection at 10% significance level. This suggests that ignoring the zero trade flows will lead to sample selection bias. Puhani

(2000) found that full information maximum likelihood estimator of the Heckman model gives better results than the two-step Heckman model. Therefore, we apply a Heckman maximum likelihood estimation by using full information of equation (14) and (15).

Table 1 shows the estimation results for the Heckman maximum likelihood model. With the parameters estimated below, we can solve for ρ, γ and θ. Note that the coefficient for the disease status term $\ln(1+\delta)$ is negative, which means that when state *i* losses its TB disease free status (δ=1), the number of cattle that are sold to another states decreases. However, the magnitude of impact of the disease status on cattle trade is small. Holding other independent variables fixed, the number cattle traded between two states decreases by $e^{(\overline{O}-\theta\ln\overline{c}-\theta\rho\ln\overline{d}+\ln 2)} - e^{(\overline{O}-\theta\ln\overline{c}-\theta\rho\ln\overline{d}_i^{\omega}+\ln 1)} = e^{\overline{O}}(\overline{c}\overline{d}_i^{\omega\rho})^{-\theta}$ = 2.6 ($\overline{O},\overline{c},\overline{d}$ are the mean values for the outward dummy, input costs, and distance, respectively), when the seller's state loses its disease free status. In addition, both distance (d_{ji}^{ω}) and input cost (c_i) have a negative impact on cattle trade.

Table 1 Maximum likelihood estimation of the Heckman model

	Coef.	Std. Err.	P value	[95% Conf. Interval]	
Outward dummy	0.33	0.25	0.191	-0.16	0.83
$\ln c_i$	-1.84	0.16	0	-2.15	-1.53
$\ln d_{ji}^{\omega}$	-0.40	0.095	0	-0.57	-0.23
$\ln(1+\delta)$	-0.73	0.66	0.271	-2.02	0.58
$\zeta_{\varepsilon\mu}$	0.82	0.043			
σ_{ε}	4.08	0.21			
Inverse Mills ratio	3.36	0.34			

bTB disease dynamics for cattle movement

The bTB disease dynamics for cattle movement are modeled using a variation of Barlow et al.'s (1998) model. The model comprises difference equations that simulate changes in numbers of three herd categories in each of 48 states. In each state, farm-level herds are divided into one of the three types: (1) herds that are healthy but susceptible (with the number in state *i* denoted S_i), including accredited herds and those becoming reaccredited after coming off movement control; (2) herds that are infected with bTB and identified as such (with the number in state *i* denoted M_i); and (3) herds that are infected with bTB but not detected (with the number in state *i* denoted I_i). We also assumed those infected and identified cattle herds will be under movement control due to government regulations. Cattle trade is assumed to occur only between herds that are either susceptible (*S*) or non-detected (*I*). The total number of herds in state *i* is given by $Z_i = S_i + I_i + M_i$.

Unlike Barlow et al.'s (1998) model, we assume the force of infection is stochastic. Specifically, we assume that each farm getting infected via cattle purchases follows a Bernoulli random process. Let η_{ji} be the probability that one cattle farm in state *j* becomes infected after a purchase of one cow from state *i*. This probability is given by $\eta_{ji} = \varphi I_i / Z_i$, where φ is the prevalence rate of a typical infected farm. If the per farm cattle imports from state *i* to state *j* is $n_{ji} = x_{ji} / Z_j$, then the probability of that one susceptible cattle farm does *not* become infected due to its trades with farms in state *i* is $\chi_{ji} = (1-\eta_{ji})^{n_{ji}}$. The probability that the farm does *not* become infected due to trades with farms in *all* states is $\prod_{i=1}^{N} \chi_{ji}$, so that the probability that the farm *does* become

infected is $\varsigma_i = (1 - \prod_{i=1}^{N} \chi_{ji})$. Note that this distribution will change over time as infection risks change and alter trade patterns, x_{ji}.

Define Ψ_j as an infection dummy variable for one susceptible farm in state *j*: Ψ_j =1 if the farm becomes infected via its cattle imports and Ψ_j = 0 otherwise. This dummy variable is a Bernoulli random variable with the probability that Ψ_j = 1 equal to ς_i. The total number of infections is determined by taking S_j draws of this variable and summing the values of those draws, i.e., the number of new infections equals $\sum_{k=1}^{S_j} \Psi_j^k$, where Ψ_j^k represents the *k*th draw of the random variable Ψ_j.

There are two other transition rates in the model. Infected but undetected farms transition are periodically tested by the government or may be detected via slaughterhouse testing. Following Barlow et al. (1998), $\xi_j = a_j + m/\tau_j$ is the rate at which "*I*" herds become detected and go on movement controls. Here, a_j is the rate of slaughterhouse detection, *m* is the test sensitivity per herd, and τ_j is the herd testing interval in state *j*. Finally, q_j is the average length of time a farm remains on movement controls, so that 1/ q_j is the rate at which the farm moves off of movement controls and transitions back to the susceptible state.

Given this specification for transition probabilities and rates, bTB disease dynamics for *N* = 48 states can be presented in the following difference equations, where the *t* subscripts are time-indices:

$$(16) \qquad S_{j,t+1} - S_{j,t} = \frac{M_{j,t}}{q_j} - \sum_{k=1}^{S_{j,t}} \Psi_{j,t}^k + \xi_j I_{j,t}$$

(17) $$I_{j,t+1} - I_{j,t} = \sum_{k=1}^{S_{j,t}} \Psi_{j,t}^{k} - \xi_{\mathrm{j}} I_{j,t}$$

(18) $$M_{j,t+1} - M_{j,t} = \xi_j I_{j,t} - \frac{M_{j,t}}{q_j}$$

Discussion

The final step is to specify parameter values for the epidemiology model and to link this model with the gravity model specified above. This final linkage produces joint behavioral and epidemiological feedbacks that are missing from existing analyses. We are in the last stages of developing this final linkage, and we will have numerical results to go along with this model by the time of the AAEA meetings in July. Still, there is value at this point of simply laying out a framework to address the limitations of current approaches.

References:

Anderson, J. E. 1979. “Theoretical foundation for the gravity equation” *American Economic Review* 69(1): 106-116.

Anderson, J. E., and E. van Wincoop. 2003. "Gravity with gravitas: A solution to the border puzzle." *American Economic Review* 93 (1): 170-192.

Anderson, J. E., and E. van Wincoop. 2004. "Trade costs." *Journal of Economic Literature* 42 (3): 691-751.

Barlow, N.D., J.M. Kean, G. Hickling, P.G. Livingstone, and A.B. Robson. 1997. "A simulation model for the spread of bovine tuberculosis within New Zealand cattle herds." *Preventive Veterinary Medicine* 32(1-2):57-75.

Barlow, N.D., J.M. Kean, N.P. Caldwell, and T.J. Ryan. 1998. "Modelling the regional dynamics and management of bovine tuberculosis in New Zealand cattle herds." Preventive Veterinary Medicine 36(1):25-38.

Bossenbroek, J. M., C.E. Kraft, J.C. Nekola, 2001. “Prediction of long-distance dispersal using gravity models: Zebra mussel invasion of inland lakes” Ecological Application 11(6): 1778-1788.

Bossenbroek, J. M. Johnson, L.E., Petters, B., Lodge, D.M.2007. "Forecasting the expansion of zebra mussels in the United States." *Conservation Biology* 21(3): 800-810.

Eaton, J., and S. Kortum. 2002. "Technology, geography, and trade." *Econometrica* 70 (5): 1741-1779.

Eichengreen, B. and D.A. Irwin, 1998. “The Role of History in Bilateral Trade Flows” Jeffrey A. Frankel, editor. pp33 - 62

Emlinger, C., F. Jacquet, and E. C. Lozza. 2008. "Tariffs and other trade costs: assessing obstacles to Mediterranean countries' access to EU-15 fruit and vegetable markets." *European Review of Agricultural Economics* 35 (4): 409-438.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error", Econometrica, 47, 153-161.

Puhani, P.A. 2000. "The Heckman correction for sample selection and its critique." *Journal of Economic Surveys* 14(1): 53-68.

Rose, A. K., and E. van Wincoop. 2001. National money as a barrier to international trade: The real case for currency union, pp. 386-390.

Samuelson, P. 1973. “Deadweight Loss in International Trade from the Profit Motive?” *Leading Issues in International Economic Policy: Essays in Honor of George N. Halm*, C. Fred Bersten and William G. Tyler, eds., Lexington, Mass., D.C. Heath and Co.

Shields, P.A. and Mathews, Jr., K. H. 2003. “*Interstate Livestock Movements*” Electronic Outlook Report from the Economic Research Service No. (LDPM10801).