Bayes Network Analysis of Trade Effects of Currency Unions and Free Trade Agreements

Phillip Baur
The Boston Consulting Group

Viktor Winschel
University of Mannheim

August 14, 2009

Abstract

We investigate the effects of trade and currency unions on trade within the reduced form gravity model. The database is a standard database in this field for around 200 countries over the last fifty years. The statistical model we use is a Bayes network from the machine learning and the artificial intelligence research which generalizes many standard econometric models. The emphasis in this paper is on a simplified specification of models with discrete and continuous variables and on simultaneous equations time series models. The results indicate that trade and currency unions have substantially lower effects on trade than derived in previous empirical work which we reproduce as special cases of our statistical framework.

JEL Classification: C25, C51, C63, F13, E42
Keywords: Bayes networks, currency unions, free trade aggrements

1 Introduction

We quantify the effects of currency unions and free trade agreements on trade. The economic framework is the gravity model. The statistical setup is a Bayes network developed in the artificial intelligence and machine learning literature. Our approach is to represent some common econometric specifications in the related work as Bayes networks and to demonstrate how easily generalized specifications can be expressed within this framework.

In the next chapter we shortly introduce Bayesian networks as they might be not very well known to most economists. In section 3 we present the basic gravity model of trade which forms the economic theory of this paper. We also discuss some shortcomings of the related work and extend the basic setup in sections 4 and 5. In the last section 6 we summarize our results.

2 Bayes Networks

The gravity model provides an interesting example where Bayes network techniques can excel. The real payoff of these techniques comes in complicated models, since Bayes networks can easily combine usual regression models, factor models, principal component techniques, mixture models, hidden markov models, state space models or regime switching models to name a few. To make this point clear: it is easy to formulate and estimate for example a model with a time series structure where in one period we have two states and a factor analytical model, at another time step ten states and some principal component structure and yet at another time step some mixture models. However, we will hopefully convince the reader that even the simple gravity model can already profit from being analyzed by Bayes networks.

Since Bayes networks are not very well known in the economic and econometric community we will first have a look on a characterization from a prominent researcher in this field, Jordan (1998):

> Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering – uncertainty and complexity – and in particular they are playing an increasingly important role in the design and analysis of machine learning algorithms. Fundamental to the idea of a graphical model is the notion of modularity – a complex system is built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to

> data. The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.
>
> Many of the classical multivariate probabilistic systems studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics are special cases of the general graphical model formalism – examples include mixture models, factor analysis, hidden Markov models, Kalman filters and Ising models. The graphical model framework provides a way to view all of these systems as instances of a common underlying formalism. This view has many advantages – in particular, specialized techniques that have been developed in one field can be transferred between research communities and exploited more widely. Moreover, the graphical model formalism provides a natural framework for the design of new systems.

Bayes networks represent joint densities as graphs. The visual representation reduces the complexity of the modelling process and elucidates the qualitative features of the assumed causal relations among the modelled variables. Graph theory is then used to automatically derive the inference algorithms tailored to the structure of the analysed model.

A Bayes network represents the joint distribution of a set of variables, e.g. $\{X, Y, Z\}$, as a directed acyclic graph combined with a set of conditional probabilities which decompose the joint of the involved variables. The nodes of the graph represent the variables and the abscence of edges represents conditional independence assumptions encoded in the conditional probabilities. For example, while $P(A, B, C) = P(A|B, C)P(B|C)P(C)$ is a complete graph with edges connecting all nodes ($B \to A, C \to A, C \to B$) a conditional independence assumption like $P(A|B, C) = P(A|B)$ leaves out the edge $C \to A$.

The usual procedure is to specify a prior $P(X)$ and a model for Y given X, i.e. $P(Y|X)$. We are then interested in the implied marginal $P(Y)$. After observing Y the inference process amounts to "reasoning against the directions of the arrows" and thus to evaluate $P(X)$ given data on Y, which is again a marginal.

The algorithmic problem is therefore to implement a multiplication $P(X, Y) = P(Y|X)P(X)$ and an addition as in $P(X) = \sum_Y P(X, Y)$ for the inference problem $P(X) = \sum_Y P(Y|X)P(X)$. For discrete variables this is all to be done. In case of continous variables we have either an analytical expression for the integral, which comes instead of the addition or in case of an approximative analysis the integral has to be discretized and results again in a sum. Essentially the problem is to decide the sequence of summations and multiplications, as $ac + bc$ needs 3 floating point operations whereas $(a + b)c$ only two.

There are many different algorithms proposed in the literature to derive the optimal sequence of additions and multiplications. A simple one, the elimation algorithm, calculates the marginal of one variable per run. Usually we need several marginals at once. This can be accomplished efficiently with the sum-product algorithm which avoids repeated calculations of the same intermediate sums and products for the different marginals. The restriction is that this algorithm is available only for trees. The junction tree algorithm calculates the marginals for general graphs by decomposing large graphs into several smaller ones, called cliques, according to the underlying structure of independence assumptions. The resulting graph is a tree where each node contains several variables. Huang and Darwiche (1996) describe the junction tree algorithm in a very intuitive way.

The statistical analysis of econometric reduced form models with Bayes networks is the first step towards decision networks where decision and utility nodes are added to the graph. The problem here is to decide on the values of the decision nodes where the utility is maximized within an uncertain setting - the usual decision theoretic model cast in a graphical representation. Decision networks in turn can be extended to graphical games where several decision makers may interact strategically. The decomposition of game networks is then not according to conditional independence assumptions in the joint density but according to the strategic independences of the payoff functions. A good starting point for decision and Bayes networks is the book by Kjaerulff and Madsen (2007).

However, we take advantage only of the basic set up of Bayes networks for reduced form econometric models. The main emphasis in our application here is on hybrid models where we mix discrete and continuous variables. Within our gravity model discrete variables D show up only as parents of continuous variables C, so conditional densities are of type $P(C|D)$ which in linear models allows for exact inference. This essentially involves a subdivision of the data sample with separate regressions. This complication is hidden from the modeller and in turn greatly simplifies the analysis. For causalities which point from continuous to discrete variables, $P(D|C)$, we would have to resort to approximative Bayes network algorithms. The second technical topic in this paper is to demonstrate how complicated simultaneous equation systems can be easily modelled with Bayes networks. This again demonstrates their power to cope with complicated models.

3 Model

The gravity model essentially explains trade by income and distance. This basic specification is usually completed by additional explanatory variables including binary intercept regression dummies to represent policy regimes such as membership in trade or currency unions.

Conclusions drawn from such setups have gained policy relevance due to the introduction of the Euro currency. Consequently, a stream of literature has evolved around specification search and modeling techniques to increase robustness and to track possible misspecifications, see for example Baldwin and Taglioni (2006). The diversity of existing estimates indicates the potential bias inherent in applied specifications, which culminates in meta studies as in Rose (2004).

3.1 Gravity Fixed Effects Model

In the trade literature Glick and Rose (2002) were among the first to use the fixed effects model via a least squares dummy variables estimator (LSDV) on a pooled time-series cross-sectional data set.

The strength of the fixed effects LSDV estimator is that it takes advantage of the time-series character as well as the cross-sectional dimension of the data delivering long-run estimators for dynamic factors. Pooling time-series data even allows for non-stationarity in individual time series.

Phillips and Moon (1999) state that even if noise is strong in individual time-series regressions, it can often be characterized as independent across sections. Consequently, the strong effect of the residuals can be attentuated via pooling while the strong signal of the regressors is retained.

Baltagi, Kao, and Liu (2008) demonstrate that the fixed effects estimator remains not only consistent but also asymptotically normally distributed in a spurious setup with non-stationary regressors and serially correlated and non-stationary disturbances for large time and cross sections. In case of cointegration the fixed effects LSDV estimator remains as efficient as a GLS estimator.

3.2 Structural Shortcomings

In spite of its robustness the panel fixed effects model via LSDV reflects particularly well the estimation bias issue. Bun and Klaassen (2007) show that the residuals of the LSDV setup of Glick and Rose (2002) exhibit trends over time and thus substantially disturb estimates for currency union membership and trade union membership. Bun and Klaassen (2007) propose to extend the standard model with a linear pair-specific time trend. As a consequence, the currency union membership dummy variable is reduced from 86 to 25 percent for the Glick and Rose (2002) data set. Bun and Klaassen (2007) propose gradual changes in transportation costs and trade liberalization at pair-specific speed to be responsible for pair-specific disturbance trends.

The changes in estimators from small changes of the model are large. And since the fixed effects gravity model is a lean model, which is forced to deal with large amounts of heterogeneity, the suspicion is that it suffers from omitted variable bias. Thus, even the Bun and Klaassen (2007) setup has most likely not yet reached sufficient precision for policy recommendations. In the following sections we will use Bayes networks to examine the bias and variation in estimates.

4 Horizontal Extension of the Model

In this section we will use a conditional linear Gaussian Bayes network to examine changes in the estimation of a currency union and trade effect in the Glick and Rose (2002) data set.[1] We will broaden the model in a horizontal fashion, meaning that we will extend the model framework and alter the variable representation.

In the first step we translate the standard panel fixed effects setup into a Gaussian Bayes network. The underlying standard model with the Bun and Klaassen (2007) extension is given by:

$$\begin{aligned}\log(TRADE_{ijt}) = \beta_1 \log(GDP_{ijt}) + \beta_2 \log(POP_{ijt}) + \delta_1 CU_{ijt} + ...\\ + \delta_2 FTA_{ijt} + \eta_{ij} + \tau_{ij} \cdot t + \lambda_t + \epsilon_{ijt}\end{aligned}$$

where η_{ij} represents the pair specific time invariant fixed effect and $\tau_{ij} \cdot t$ the pair specific linear time trend. The λ_t term is a macroeconomic binary dummy for every year and ϵ_{ijt} the disturbances. CU[2] stands for the binary currency union dummy variable, FTA for the free trade agreement membership. $TRADE$, GDP and POP represent the time variant variables bilateral trade, national income and population, respectively, aggregated for countries i and j at time t.[3]

[1] We reduce the data set by dropping all time-series shorter than 15 observations.

[2] In the Glick and Rose data set currency unions are transitive thus if two countries have a fixed exchange rate with a common currency they are considered to be in a currency union as well.

[3] Usually $GDPPC$ is used instead of POP, however since $GDPPC$ is calculated by GDP/POP both POP and $GDPPC$ result in the same explanatory power since all residual changes at the side of the GDP regressor stem from POP in GDP/POP.

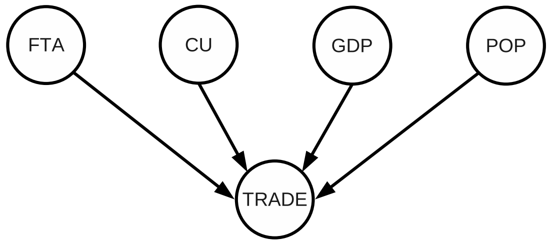

Figure 1: DAG Specification 1

The specification can be translated into a directed acylic graph in Figure 1 with nodes N and edges E:

$$N = \{TRADE, GDP, POP, CU, FTA\}$$
$$E = \{(GDP, TRADE), (POP, TRADE), (CU, TRADE), (FTA, TRADE)\}.$$

Every node represents a continuous random variable that follows linear Gaussian probability density function. The arrows encode the causal assumption of the underlying a linear regression. In the graph the regressor nodes $\{GDP, POP, CU, FTA\}$ are the parents of the regressand $TRADE$.

The terms $\eta_{ij} + \tau_{ij} \cdot t + \lambda_t$ are not represented in the graph to reduce complexity, since there are several thousands of these variables for the Glick and Rose (2002) data set. However, there are graphical devices called plates to represent repeating graph structures in cross or times series dimension.

We applied the Frisch-Waugh theorem on partitioned models and first regressed $\eta_{ij} + \tau_{ij} \cdot t + \lambda_t$ on the remaining regressors and the regressand. Thus, the network nodes stand for the deviations from their unit means which are the disturbances of the first stage regressions.

After the network has been specified we estimate the parameters of every node from the Glick and Rose (2002) data set. Then, the junction tree algorithm is used to draw inference in the network and to retain the marginal distribution of $TRADE$ given realizations of all four combinations of the two intercept dummy nodes $\{CU, FTA\}$.[4]

The marginal distribution of the child node corresponds to the predicted value in a linear regression. Table 1 shows predicted values for the Glick and Rose (2002) setup without country pair time trends and the Bun and Klaasen (2007) setup with pair time trends included. Including pair time trends sweeps out the trended residuals' upward bias in the estimators. The

[4]Beta coefficients that link all parents (regressors) with the child (regressand) are identical to standard OLS and inherent in the conditional probability distribution of the child node.

Table 1: Inference Results - Trade Volumes

Description	Marginal Distribution of Trade Node (σ)			
	No Pair Trend	With Pair Trend		
Given Evidence	Spec.1	Spec. 1	Spec. 2	Spec. 3
No CU no FTA	.240 (1.44)	.003 (0.96)	-.001 (0.97)	-.002 (0.98)
CU Member	.697 (1.44)	.239 (0.96)	.035 (0.75)	.040 (0.76)
FTA Member	.872 (1.44)	.314 (0.96)	.052 (0.33)	.053 (0.35)
CU&FTA Member	1.568 (1.44)	.552 (0.96)	.008 (0.33)	.004 (0.29)
No. Obs.	192831	192831	192831	192831
No.Pairs	6022	6022	6022	6022
BIC (Partition Exc.)	-299288	-265734	-265182	-265182
BIC (Partition Inc.)	-336223	-339305	-338753	-338753

Notes: Results display inference results conditional on different realizations of policy regimes drawn from a Bayes Network with data already partitioned in the first step on year dummies, fixed effects and partly also pair time trends (see Table). Thus, trade volumes should not be compared directly between regression, since different partitions are applied, with and without pair time trends which alter the absolute height of trade volumes. Instead one should compare differences inside the regressions for different policy regimes with each other.

Glick and Rose (2002) setup predicts an increase in trade mean through membership in a currency union of $.697 - .24 \approx 46$ percent. For the Bun and Klaassen (2007) specification this increase shrinks down to $.239 - .003 \approx 24$ percent. The effect of joining a free trade agreement is reduced from $.872 - .24 \approx 63\%$ to $.314 - .003 \approx 31$ percent increase.[5]

We immediately observe that the Bayesian Information Criterion (BIC) reveals a better fitness of the Bun and Klaassen (2007) model. However, since time trends, fixed effects and pair time trends are leveled out in the first stage regression, we have to correct the Bayesian Information Criterion for the additional parameters. For the model with pair time trends this implies a set of 6021 additional regressors. After correction, the Bun and Klaassen BIC drops below the Glick and Rose model. However, these values

[5]It has to be noted that the estimator is unbiased and consistent, but not efficient, since residuals prove to be strongly serially correlated - including time trends has not changed this fact. We follow Wooldridge (2002) to test for panel serial-correlation. Bun and Klaassen (2007) compute standard errors that are robust to heteroskedasticity and serial as well as cross-sectional correlation. We dispense with that since we are at this point not interested in precise inference but in showing changes in the mean level of estimates due to misspecification.

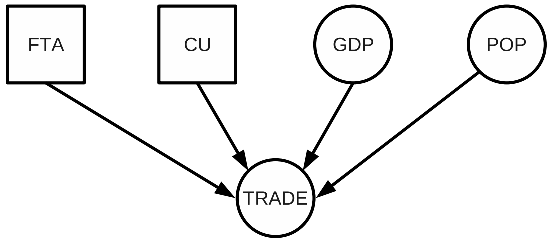

Figure 2: DAG Specification 2

do not form a basis for choosing between both specifications, since the Glick and Rose estimation suffers from omitted variable bias.[6]

4.1 Representation of discrete variables

Translating a linear regression into the graphical language of Bayes networks elucidates the doubtful nature of intercept dummy variables. Discrete intercept dummies are treated as continuous variables.

Moreover, the coefficients are interpreted as if the policies can be switched on and off which is based on the assumption that there exists only a partial effect. It is doubtful whether a ceteris paribus analysis makes sense for policy variables. If the analysis claims to be policy relevant it should incorporate additional channels of the effects such as level differences in other explanatory variables. This and an intuitive and natural way to represent discrete policy variables in Bayes networks is the emphasis of the following analysis

In the framework of a conditional linear Gaussian Bayes network discrete and continuous variables can be clearly distinguished. In Specification 2 in Figure 2 the two nodes $\{CU, FTA\}$ are no longer treated as continuous but as explicitly discrete, represented by their rectangular shapes. The underlying conditional linear Gaussian distribution in the child node of interest, see Kjaerulff and Madsen (2007), is now:

$$L(Y|X = x, I = i) = N(A_i + B_i^T x, C_i).$$

The child nodes distribution is no longer only conditioned on its continuous parents, but also on the possible states i of the discrete parent nodes. The usuall econometric notion for this specification is a full interaction variable.

The model has now continuous $X = \{GDP, POP\}$ and discrete $I = \{CU, FTA\}$ variables. Practically, a separate regression is run for every possible combination of the discrete states, in the case of two discrete nodes

[6]For a graphical traction and a more detailed analysis of the upward bias in the estimators see Bun and Klaassen (2007)

with each two realizations (member and no member) this implies running four different regressions to specify the conditional probability distribution for the $TRADE$ node given its parents.

By that we take into account the fact, that different states of policies might not only have marginal effects on the level of our linear continuous relationships but also change the continuous relationships themself and the mean level of other explanatory variables. Membership in a trade union might for example come along with different mean incomes and different elasticities concerning trade.

Estimating Specification 2 on the second stage data, freed from fixed effects, time effects and pair-time trends, results in a higher BIC value than the Bun and Klaassen and Glick and Rose specifications. This confirms the hypothesis that membership in currency unions and free trade agreements have not only a partial effect on trade but also effects other explanatory variables relationships with trade. Considering these structural differences the effect of currency union membership on trade decreases to $.035 + .001 \approx 3.7$ percent, while membership in free trade unions rises trade levels by $.051 + .001 \approx 5.5$ percent. The predictions are no longer restricted to the partial effect of the policy changes but incorporate changes in income and population means and changes in their relationships with trade.

This brings along another interesting result. Conclusions drawn from the addition of several partial effects can be delusive. With structural states represented through intercept dummies one would conclude that joining a CU and an FTA would result in the aggregate trade effect of both intercept coefficients. The analysis above shows that the combination of both policies falls indeed below the individual membership levels, which indicates that the combined policy has to be considered as a separate policy regime, if one does not want to be trapped by false conclusions of their combined effects.

4.2 Multiple Equations

The absence of arrows in between the regressor nodes underlines one foundation of the simple linear regression model. Regressors are assumed to be independent.[7] By not placing an arrow in between two regressors, the correlation between two variables is not interpreted as a causal link and possible interaction is discarded from the analysis. As long as a ceteris paribus effect is explicitly intended by the researcher this does not pose a problem. However, it seems that empirical research restricts its scope to single equation

[7] It is important to notice that this holds only for their partial effects and not the variable values overall.

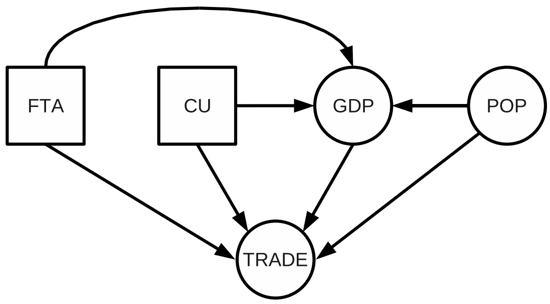

Figure 3: DAG Specification 3

analysis out of convenience and might thus neglect multidimensional aspects.

Policy relevant modeling should thereby not restrict itself to partial ceteris paribus effects but derive the composite effect of all relevant variables on the target variable. This consequently motivates to model not only direct causal effects on trade, but also causal effects among the regressors of the original setup, to give a more precise estimate of trade changes under different policy regimes. Bayes Networks easily allow adding potential causal relationships and evaluating corresponding BICs for the resulting graphs.[8]

The DAG Specification 3 in Figure 3 outperforms a set of 11 potential graphs, that could explain effects among our 5 variables, obtained from local search over possible specifications. POP has a causal effect on GDP and the policy states FTA and CU directly influence GDP but not POP in the dominant specification. This is intuitive considering trade stimuli on GDP for open economies and the extension of markets. Independence of POP from FTA and CU seems as well intuitive.

Population growth will not be affected by policy regimes directly and the effect of GDP on POP is dominated in size by the reverse impact of POP on GDP. Stepping from Specification 2 to 3 causes only slight changes in predicted trade values. Nevertheless, taking into account effects of the policy variables on POP and GDP slightly increases the mean predicted value for $TRADE$ under CU and FTA memberships respectively (see Table 1).

4.3 Dynamics

Several authors have applied dynamic models to tackle the problem of serial correlation directly via lags and not indirectly through robust standard deviations. Bun and Klaassen (2007) apply an autoregressive distributed lag

[8] This procedures allow the search for the direction of causality up to Markov equivalence, see e.g. Murphy (2001).

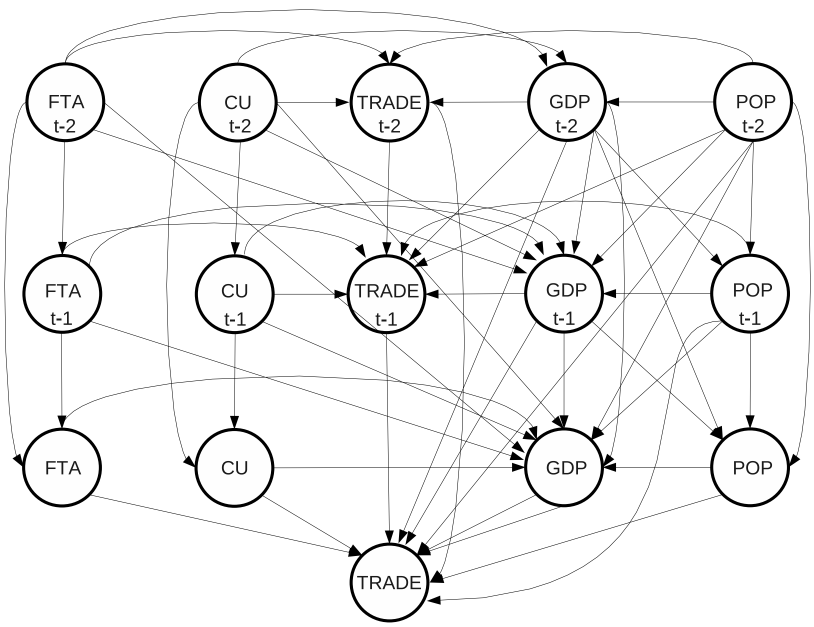

Figure 4: DAG Specification 4

model, De Nardis, De Santis, and Vicarelli (2008) a system GMM dynamic panel data estimator.

Including lags into a fixed effects setup is problematic since mean transformation causes correlation of the lagged regressand as regressor and the error term leading to a biased but consistent estimator gives no serial correlation in the residuals. However, this finite sample bias diminishes with increasing time spans. Judson and Owen (1997) show that the LSDV estimator bias for pooled panels with time series of 30 periods exhibits a relatively small bias of 3 to 20 percent.

Since our time series vary between a length of 15 to 47 years, we believe this bias not to be substantial for our data set. We have added two lags for every variable for $\{TRADE, GDP, POP, CU, FTA\}$ and searched over 13 different structures to find the BIC dominant one that could causally explain interactions of our variables.[9] Figure 4 displays the BIC maximal specification. The pattern consists of three static gravity regressions (like in Specification 3) that are dynamically combined. It is defined by the following

[9]Bun and Klaassen (2007) have tested that the inclusion of two lags is sufficient to sweep away serial correlation in the residuals.

conditional and marginal Gaussian probability distributions:

$$P(T_t|T_{t-1}, T_{t-2}, GDP_t, GDP_{t-1}, GDP_{t-2}, POP_t, POP_{t-1}, POP_{t-2}, ...$$
$$CU_t, FTA_t)$$
$$P(T_{t-1}|T_{t-2}, GDP_{t-1}, GDP_{t-2}, POP_{t-1}, POP_{t-2}, CU_{t-1}, FTA_{t-1})$$
$$P(T_{t-2}|GDP_{t-2}, POP_{t-2}, CU_{t-2}, FTA_{t-2})$$
$$P(GDP_t|GDP_{t-1}, GDP_{t-2}, POP_t, POP_{t-1}, POP_{t-2}, ...$$
$$CU_t, FTA_t, CU_{t-1}, FTA_{t-1}, CU_{t-2}, FTA_{t-2})$$
$$P(GDP_{t-1}|GDP_{t-2}, POP_{t-1}, POP_{t-2}, CU_{t-1}, FTA_{t-1}, CU_{t-2}, FTA_{t-2})$$
$$P(GDP_{t-2}|POP_{t-2}, CU_{t-2}, FTA_{t-2})$$
$$P(POP_t|POP_{t-1}, POP_{t-2}, GDP_{t-1}, GDP_{t-2})$$
$$P(POP_{t-1}|POP_{t-2}, GDP_{t-2})$$
$$P(POP_{t-2})$$
$$P(CU_t|CU_{t-1}, CU_{t-2})$$
$$P(CU_{t-1}|CU_{t-2})$$
$$P(CU_{t-2})$$
$$P(FTA_t|FTA_{t-1}, FTA_{t-2})$$
$$P(FTA_{t-1}|FTA_{t-2})$$
$$P(FTA_{t-2})$$

Figure 5 shows graphically the inference results for different policy states and different combinations of the lagged structural variables on the mean predicted trade for the estimated DAG. With two lags it is possible to track the dynamics of joining a CU or a FTA over three periods as well as of leaving a currency or trade union.

The figure shows that joining a CU, while being no member in a FTA, comes along with about 2.8 percent predicted growth in the first period of membership, compared to no membership volumes, followed by 3.3 percent for the second period and a lower 1.7 percent growth for the long run membership (three or more periods inside a currency union).

Leaving a currency union causes predicted drops similar to the increase pattern, while the first year outside comes along with the steepest drop.

Joining a FTA stimulates trade more than joining a CU. Trade grows stronger and final predicted trade volumes lie about 5 percent above the ones for currency union members. Countries that become members of both CU and FTA lie below trade levels of single FTA or CU members.

A dynamic representation does not only permit the analysis of dynamics but demonstrates how a temporal differentiation alters structural effects. It

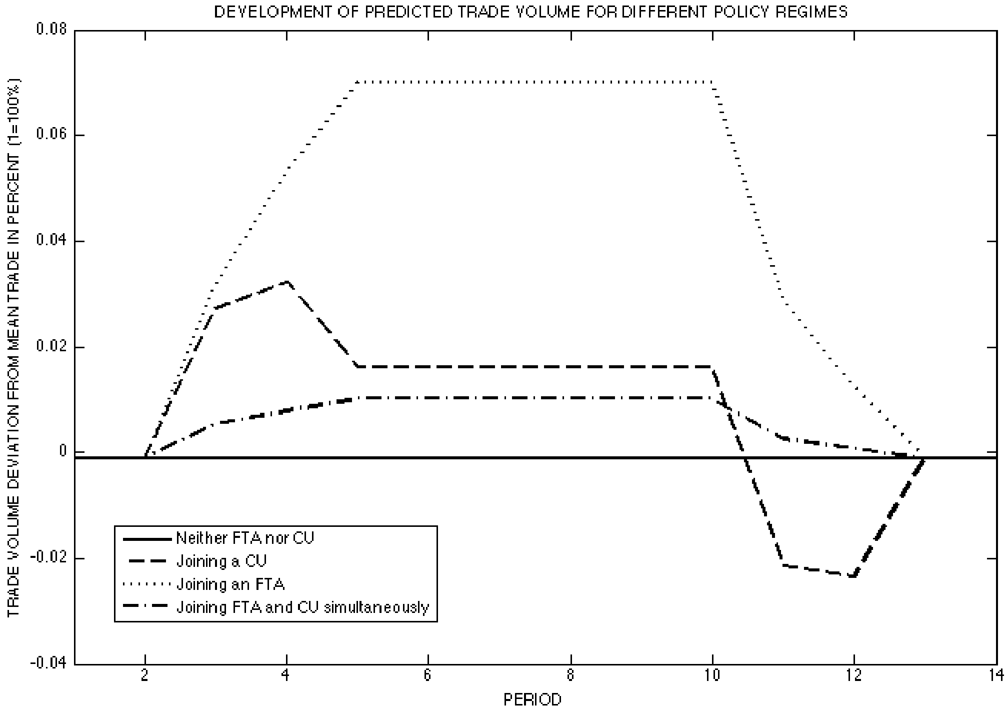

Figure 5: Policy Regimes and Predicted Mean Trade - Dynamic Tracking

allows to split entrance phase effects from long-run effects, which is not possible for static models, since they blend temporal developments into one effect. Doing so, predicted long-run trade levels for CU membership exhibit with 1.7 percent lower values compared to the static Specification 3. Predicted trade levels for FTA membership have increased from 5 to about 7 percent growth.

The specification search above can only be seen as the beginning to find a dynamic representation for the prediction of policy effects on trade. Time spans should be extended to further differentiate temporal development. Moreover, bias through included lags, eventual remaining serial correlation, and heteroskedasticity spoil the efficiency of the multi equation system constructed via Bayes Network and might even lead to slight bias and inconsitency in parts of the multiple equation system.[10] The next chapter will provide one possible solution for obtaining an unbiased, consistent and effi-

[10]Even if serial correlation in the residuals is eliminated for the present trade node, it certainly persists for the regressions in the lagged trade nodes since their static regressions individually seen contain less than two lags up to no lags at all.

cient estimator.

5 Vertical Extension of the Model

Horizontal broadening, the incorporation of pair time trends, the transformation of policy regime representation from continuous to discrete and additional causal links both in the static and in the dynamic models have caused large changes in policy regime impact the estimates. The source for such variation lies in the lean setup of fixed effects gravity model estimators that explain changes in large pools of structurally very diversive panel observations with a relatively small number of variables.

This diversity leaves footprints in the disturbances. Panel serial correlation can be detected in the estimators of the residuals with and without pair trends as well as traces of heteroskedasticity.[11] Especially serial correlation can be seen as an indicator for misspecification, rendering OLS-estimates possibly biased and certainly inefficient.

Bun and Klaassen (2007) applied Driscoll-Kraay-Newey-West robust standard errors. This allows for efficient inference but is no remedy against bias if serial correlation and heteroskedasticity are symptoms of an omitted structure. The introduction of lagged variables is more appropriate which is also proposed by Bun and Klaassen (2007). In this context it is important to absorb all serial correlation, since the combination of serial correlation and lags leads inevitably to inconsitent and biased estimators.

Additionally, serial correlation can be interpreted as a symptom of a spurious setup. $TRADE$, POP as well as GDP are non-stationary time-series.[12] Here Bun and Klaassen (2007) apply a dynamic OLS to integrate leads and lags in order to model cointegration among the two variables GDP and POP.

Instead of trying to cure ill results or mend them via lags we want to construct an estimator that is not at all or at least less infected in the first place. This approach is based on the logic, that the entire model should be refined, if it is unable to incorporate significant parts of the structure in the data, which is expressed through serial correlation.

We apply two measures. First, we want to allow for a maximum degree of variation in the estimators. We run individual regressions with $\{GDP, POP, CU, FTA\}$ on $TRADE$ for each of the 6022 country pairs with panels from 15 observations and more. Second, we exclude time series, which

[11] We applied Wooldrigde (2002) test for serial correlation in panels and a standard test for heteroskedsticity.

[12] We followed Hadri (2002) on parts of the panel to test for stationarity.

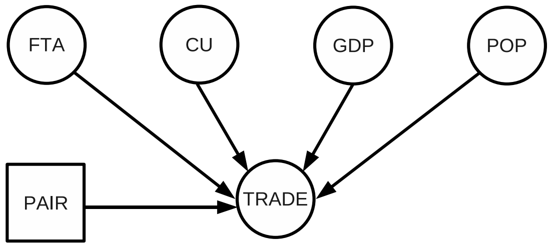

Figure 6: DAG Specification 6

can not be explained by the model. This means, we test $\{GDP, POP\}$ on stationarity via the Phillips Perron test and the residuals on ARCH effects via a standard ARCH effects test, stationarity via Phillips Perron test and serial correlation via Breusch-Godfrey test for up to 3 lags.[13] Then, we drop all time series pairs that exhibit either non-stationarity in one variable, serial correlation or heteroskedasticity. For the remaining country pair regressions we can exclude the presence of spuriousness or substantial misspecification and retain efficient estimations.

Indeed, only 41 of 6022 pairs are stationary in $\{GDP, POP\}$ as well as in the residuals and exhibit additionally no serial correlation. These results imply a very low fitness of the gravity model on an individual basis and observations are far too few to construct an estimator from this reduced data set.

Examining first differenced data in the same manner results in 2,276 country pairs that suit all criteria: stationarity, no serial correlation in individual regression and additionally no ARCH effects.[14] Consequently, 2,276 different unbiased, consistent and even efficient sets of estimators are available.

The Bayes network methodology permits a simple approach to combine these estimators and to obtain an estimation that fits the regressors on a pair basis and helps to reach the highest degree of flexibility. This can be regarded as a vertical extension of the model since we split up the estimation. In Figure 6 a discrete $PAIR$ node, whose probability table consists of 2,276 states for every country pair in the reduced sample, is included. The DAG estimated provides estimators for $\{GDP, POP, CU, FTA\}$ for each realization of $PAIR$, thus on an individual pair basis.

Then, the junction tree algorithm allows to draw inference over all pairs. A Gaussian mixture is calculated over all predicted values weighted with

[13]All tests are evaluated at 10 percent significance level; for the Phillips Perron test we include an additional time trend.

[14]We no longer use linear pair time trends - neither with the differenced data, nor with the Phillips Perron test.

Table 2: Inference Results - Growth Rates (1. Differenced Data)

Description	Marginal Distribution of Trade Node (σ)			
	With Pair Trend			
Given Evidence	Spec.1	Spec. 6	Spec. 7	Spec. 8
No CU no FTA	-.011 (1.27)	.038 (5.80)	.000 (0.67)	.033 (17.85)
CU Member	-.023 (1.27)	.037 (5.80)	-.003 (0.67)	.032 (17.85)
FTA Member	-.018 (1.27)	.034 (5.80)	-.007 (0.67)	.034 (17.85)
CU&FTA Member	-.041 (1.27)	.037 (5.80)	-.010 (0.67)	.033 (17.85)
No. Obs.	67601	67601	49083	49083
No.Pairs	2276	2276	1349	1349
BIC (Partition Exc.)	-102507	-138602	-54819	-113991
BIC (Partition Inc.)	-115425	-138602	-62347	-113991

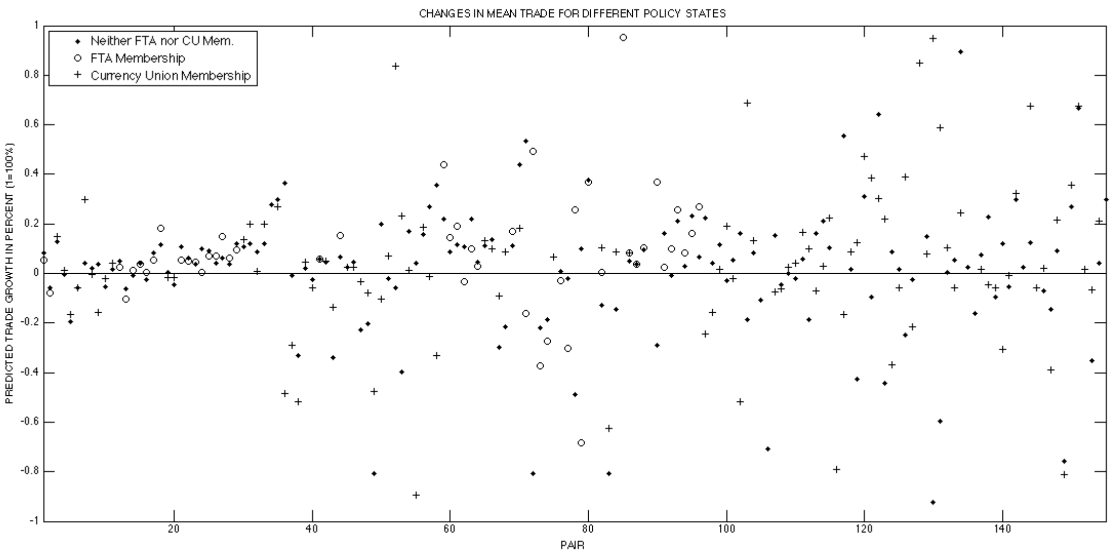

Figure 7: Policy Regimes and Predicted Mean Trade - Specification 5

the probability distribution of the discrete $PAIR$ variable. At this point, the efficient nature of the Bayes Networks methodology becomes clear, since 2,276 estimations can be conveniently combined and weighted.

The predicted mean trade differs from the one estimated via fixed effects estimator on the same reduced data set. Specification 5 predicts an average 3 percent trade growth, compared with countries neither in CU or FTA, while members of currency unions exhibit 0.1 percent smaller growth rates. Free trade agreements do not lie above sample average either. These results speak against a currency union effect on average trade growth rates. They rather suggest that trade differences in means for undifferenced data are due to level effects or misspecifcation.

Figure 7 displays difference in intercepts for currency union membership and free trade agreement membership compared to the periods with no membership for all country pairs that were or became members of currency unions or free trade agreements during their periods of measurement. Only 54 percent of all pairs that shared a currency union with each other exhibit higher trade growth during their period of membership than during the period without membership, 48 percent experience higher trade growth during their periods of free trade agreement existence. This accentuates that neither FTAs nor CUs guarantee higher trade growth. Thus, positive fixed effects results in trade levels seem to root rather in level shifts or trend absorption but not in lasting improvements in trade growth rates.

The Gaussian mixture distribution comes along with a large standard deviation which underlines the large diversity in regression coefficients once they are allowed to vary for each time series. Similar to the inclusion of country pair effects the extension can not be legitimated over a BIC comparison. A large increase in the likelihood through pair estimations can not compensate the increase in estimated parameters from 1,405 to 1,833,291.

5.1 Dynamic Estimation

In Specifications 7 and 8 we extend Specifications 1 and 6 with two lagged variables of $TRADE, GDP, POP$. Again we drop all pairs whose time series do not pass stationarity tests in the regressor variables and the residuals, serial correlation and heteroskedasticity tests on a individual pair regression basis. Moreover, we drop all pairs below 30 observations to minimize the lag bias.

The marginal distribution for predicted trade in Specification 7 obtained from the remaining 1349 country pairs shows again no increase in trade growth similar to the results of Specification 5. Instead, FTA membership shows a slight increase in predicted trade growth. Individual policy regime intercept analysis (Figure 10) does not change significantly compared to Specification 5 (Figure 7). 52 percent of all CU periods exhibit higher mean trade growth than under respective non membership periods, 48 percent for FTA membership.

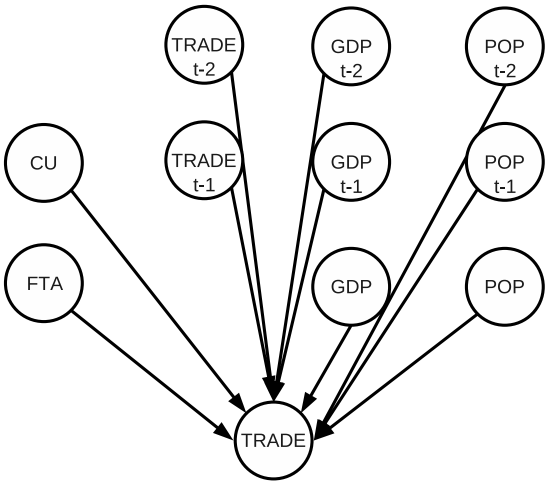

Figure 8: DAG Specification 7

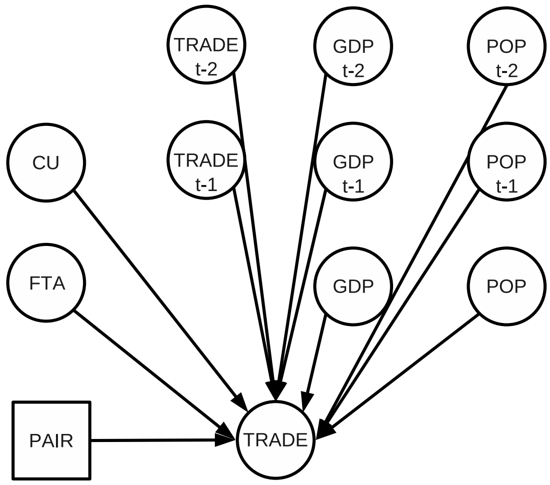

Figure 9: DAG Specification 8

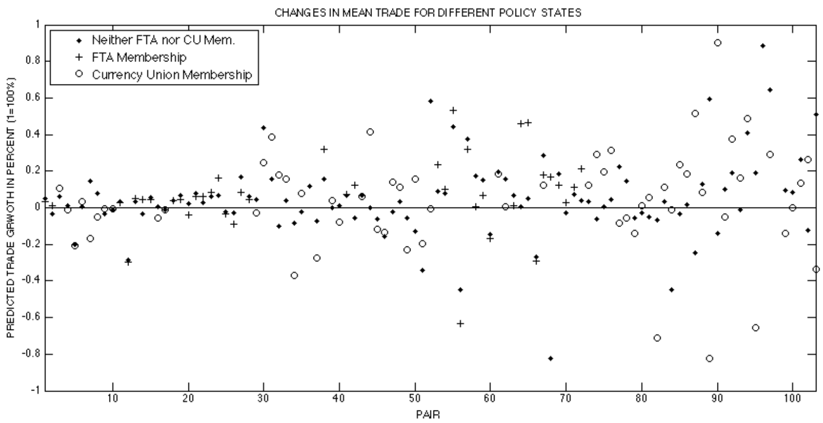

Figure 10: Policy Regimes and Predicted Mean Trade - Specification 6

Consequently we come to the conclusion that the integration of lagged variables does not alter the estimated policy regimes remarkably. However, it raises the explanatory power of the estimation, what a glance at BIC values for Specifications 6 and 8 reveals.

6 Conclusion

The horizontal and vertical extension of the fixed effects estimator via Bayes Networks has once again confirmed the fragile nature of policy relevant conclusions drawn from fixed effects gravity models. The representation of policy regimes through discrete variables and the extension of the model into a dynamic setup have substantially lowered currency and trade union effects. Moreover, these measures have shown that different combinations of policies can not be deduced from the addition of partial effects but have to be modeled separately.

Serial correlation remains a major issue in fixed effects estimation on large heterogenous data sets. The vertical extension has proven to add flexibility to estimation, decrease serial correlation and in addition to that it has also reavealed the flexibility of Bayes Networks in modeling and combining different structural regimes at the same time. Bayes networks' graphical intuition and the possibility to draw inference from complex multiple equation systems add to the picture.

Thus, Bayes networks provide a powerful toolbox for econometrics and the search for causalities. Future development should focus on the integration of issues of econometric relevance, such as the inclusion of further misspecification tests and remedies for serial correlation, non-stationarity and heteroskedasticity into the computational setups of Bayes Networks.

References

BALDWIN, R., AND D. TAGLIONI (2006): "Gravity for Dummies and Dummies for Gravity Equations," *NBER Working Paper 12516*.

BALTAGI, B. H., C. KAO, AND L. LIU (2008): "Asymptotic Properties of Estimators for the Linear Panel Regression Model with Individual Effects and Serially Correlated Errors: The Case of Stationary and Non-Stationary Regressors and Residuals," *The Econometrics Journal*, 11(3).

BUN, M. J., AND F. J. KLAASSEN (2007): "The Euro Effect on Trade is not as Large as Commonly Thought," *Oxford Bulletin of Economics and Statistics*, 69(4), 473–496.

DE NARDIS, S., R. DE SANTIS, AND C. VICARELLI (2008): "The Single Currency's Effects on Eurozone Sectoral Trade: Winners and Losers?," *Economics - The Open-Access, Open-Assessment E-Journal*, 2(17).

GLICK, R., AND A. K. ROSE (2002): "Does a Currency Union affect Trade? The Time Series Evidence," *European Economic Review*, 46(6), 1125–1151.

HADRI, K. (2002): "Testing for Stationarity in Heterogenous Panel Data," *Econometrics Journal*, 3(2), 148–161.

HUANG, C., AND A. DARWICHE (1996): "Inference in belief networks: A procedural guide," *International Journal of Approximate Reasoning*, 15(3), 225–263.

JORDAN, M. (1998): *Learning in graphical models*. Kluwer Academic Publishers.

JUDSON, R. A., AND A. L. OWEN (1997): "Estimating dynamic panel data models: a practical guide for macroeconomists," Finance and Economics Discussion Series 1997-3, Board of Governors of the Federal Reserve System (U.S.).

KJAERULFF, U. B., AND A. L. MADSEN (2007): *Probabilistic Networks for Practitioners- A Guide to Construction and Analysis of Bayesian Networks and Influence Diagrams*. Springer.

MURPHY, K. P. (2001): "Active Learning of Causal Bayes Net Structure," Discussion paper, "".

PHILLIPS, P. C., AND H. R. MOON (1999): "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67(5), 1057–1111.

Rose, A. K. (2004): "A Meta-Analysis of the Effects of Common Currencies on International Trade," *NBER Working Paper 10373*.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.