AN ECONOMETRIC MODEL OF CGE SIMULATIONS

Paper to be presented at the XIth EAAE Congress

August 23-27, 2005

Abstract

CGE models are widely used tools for economic assessments of trade policy changes. However, overall confidence in their results tends to be low. Easily accessible methods for quantitative comparisons of CGE- based simulation results do not exist. We compile a dataset of twelve recent Doha simulations and fit a linear regression model that explains the variance between simulation results on the regional level as a function of various modeling characteristics. The estimates are broadly in range with documented qualitative knowledge about modeling assumptions. The size of the sample limits general conclusions; however, an ongoing research project will extend the approach to a larger sample.

Keywords: *CGE, systematic review, Doha round, evaluation*
JEL classification: C20, C68, D58, F10, C99

1. Introduction and Research Question

Trade liberalization is one of the most promising and at the same time one of the most challenging tasks for a globalizing world. The World Trade Organization has emerged as an institution that bundles interests and arguments of countries and regions within the ongoing process of trade negotiations (WTO 2004). Economists have continuously stressed the potential gains in economic welfare that would result from liberalized markets; on the other hand, a large public especially in developed countries is highly concerned about „non- monetary" aspects of further global market liberalization (Aisbett 2004). This public watches ongoing trade negotiations critically and sometimes allies with supporters of protectionist policies.

Liberalization of agricultural markets is among the most critical issues of trade negotiations and has often been at the center of discussions. An important task for applied and agricultural economists around the world is to provide policymakers as well as a concerned public with quantitative assessments on the regional distribution of potential gains and losses from agricultural and other market liberalization.

Computable General Equilibrium models (CGE, also labeled as Applied General Equilibrium models, AGE) are state of the art for the assessment of international trade policy changes; their results sometimes enter directly into trade negotiations. CGE models have evolved as powerful tools to assess hypothetical as well as actual policy changes within a real world data framework (van Tongeren et al. 2001; Hertel 2002). However, CGE models have frequently been criticized both by economists who develop these models as well as by policymakers who focus on simulation results. Perhaps the most often stated criticisms is the so called "black box" nature of many simulation experiments. This criticism addresses the fact that the causality between assumptions entering into a CGE model and output produced by the model often remains hidden from all but those who have designed the model (Hertel 1999). Closely related to the "black box" criticism is the phenomenon that different models often produce widely differing results for identical policy shocks.

From an economic perspective, scientists analyzing trade policy changes through CGE models are suppliers of information. This information is used by a heterogeneous demand side as an input to political discussions and trade negotiations. Since this demand side for CGE- based information has usually significantly less insight into the nature of a CGE experiment than the information supplying economists have, asymmetric information will limit the efficient use of information generated from CGE models (Akerlof 1970). Because the demand side cannot judge on the quality of a specific CGE result, credibility and reliability will be placed on any simulation according to the estimated average quality of all CGE results. According to the "black box" criticism, the average confidence in CGE-based results will tend to be low. Due to this, inefficiencies arise that may in its worst directly lower the efficiency of trade negotiations or related discussions. One possible solution to the asymmetric information problem is the creation of transparency within markets; one way to create transparency is the provision of information, a different approach is the introduction of quality signaling. Information nowadays tends to be available for most contemporary models and databases; however, even few scientists usually manage to gain command of more than one or two modeling frameworks. In other words, the provision of all modeling details is a necessary condition for transparency, yet not a sufficient remedy for the asymmetric information problem, given the limited capacity on the demand side for CGE- based information to include most of economic modeling science.

From the perspective of this demand side, two important components will determine the estimated quality of simulation results: First, the question on how precise a simulation reflects reality, and second, people will ask how the simulation compares to other, related simulations. Concerning the first quality criteria, CGEs are known to do reasonably well on average, assuming normal conditions (Gehlhar 1997; Nölle and Banse 2001; Kehoe 2002). Concerning the second quality criteria, the "black box" criticism remains unsettled. Presently, concerning the "black box" criticism, no quantitative method exists that enables the comparison of CGE- based simulation results in a way that would significantly reduce asymmetric information. So far, the only solution is a detailed assessment of the complete experimental setting, due to technical and theoretical complexity for most people on the "demand side" for CGE- results not a feasible option.

Instead, various other approaches exist: First, qualitative reviews summarize the CGE- based literature with regard to a specific question. Often, authors of these reviews establish tables or frameworks where simulation results are presented in relation to various modeling features (Francois 2000; Scollay and Gilbert 2000; Hertel 2002). Secondly, increased effort has been put into the development of sensitivity analysis. Through this method, authors test the sensitivity of their results due to variation of key parameters (for instance, the Armington elasticity) and sometimes establish confidence intervals for their simulation results (Dawkins 1997; Hertel et al. 2004). A third approach to more transparency within CGE results is the synchronization and merging of models (Bertelsmeier et al. 2003), although this approach may either increase transparency as well as even decrease it through a growing complexity of such multi- model systems.

This paper introduces an experiment: the variance between simulation results from different CGE- based policy assessments is explored econometrically. Thus, this experiment proposes a new method to compare the results of CGE- based trade policy assessments in a quantitative way. The approach combines the comparative advantages of CGE models with the causality explaining methods of econometric estimation. Although not well established yet, the approach may lead towards the development of sound criteria for a reliable comparison of CGE results and thus might help to reduce asymmetric information concerning information on the effects of trade policy changes.

This paper is organized as follows: in the next section we introduce an econometric model that explains simulated welfare changes as a function of applied policy shocks and various other factors that may influence simulation results.

Section 3 presents a sample of 12 comparable assessments of trade policy changes under the WTO "Doha" round. Out of this sample, we compile a dataset; next, we fit a linear regression model to this dataset. Section 4 presents results of this estimation; Section 5 discusses our findings according to further research questions and implications for policy making.

2. Methods: An Econometric Model of Simulated Welfare Effects

In general, CGE models link economic theory to observed accounting data from regions and countries in order to measure the changes that occur in the data after certain policy variables within the model have been shocked. CGE models thus allow for experimental settings with hypothetical policy scenarios; these settings abstract from reality through assumptions that are common, if not fundamental, in economic theory. The way in which these theoretical assumptions are implemented into a model determine those aspects of reality that a specific model can highlight for closer analysis. Per definition, no model captures all of reality. Instead, scientific models must be suited for the question they are employed to answer (Heisenberg 2000).

Concerning the interpretation of simulation results form CGE models, this implies that an understanding of the entire experimental setting is necessary. These settings will determine simulation results and thus, these results cannot reasonably be interpreted without their experimental context. In addition, factors that are exogenous to the experimental setting may as well be influential to simulation results. Equation (1) presents a general model of CGE- based simulation results:

$$Y = f\,(X,M,S) \tag{1}$$

With Y = simulated change of an endogenous variable, such as economic welfare measured as equivalent variation (EV); X is a vector of policy modifications ("shocks") to the model; M is a vector of model specifications such as behavioral assumptions, data specifications and exogenous parameters; S is a vector of factors that influence the experimental settings exogenously, for example the technical progress of computing capacities.

Based on these assumptions, our experiment looks as follows: Out of a sample of CGE- based simulations we compile a dataset which captures the experimental settings of these simulations in a numerical way. Next, we fit a linear regression model to the data in order to test the hypothesis whether simulation results can be explained by the observed variables in a quantitative way. Equation (2) presents our linear model to be estimated:

$$y_{ijr} = \beta_0 + \beta\, x_{ij} + u_{ijr} \tag{2}$$

With: r = an index of simulated welfare changes observed at the regional level; x = a vector of explanatory variables that account for various assumptions of the CGE models in focus;
i = publication i for I publications, i= 1,…, n;
j = modelling framework (e.g. GTAP, BDS), j= 1,…, n,
β a vector of coefficients to be estimated;
u is the error term, for u_{ijr} we assume N~(0, σ^2).

3. Data: A sample of recent "Doha round" Simulations

In this section we employ a survey of recent CGE- based policy assessments in order to establish a dataset which can be accessed econometrically. The sample has been compiled by the authors of UNCTAD (2003) with the intention to support a political discussion. We have chosen this sample because it clearly targets the "demand side" of CGE- based assessments. At the same time, the sample is rather small and thus easy to handle for our experimental purpose.

The goal of the effort by UNCTAD *"... is to go back to basics..."* (UNCTAD 2003), as the authors state in their preface; they survey various CGE- based trade policy assessments and reach the conclusion that developing countries may still gain from trade liberalization as long as they manage to play an active role in the upcoming negotiations. For this purpose UNCTAD (2003) has published a summary and discussion of twelve quantitative policy assessments of potential scenarios under the "Doha" round of trade negotiations. The publications in focus are (Anderson et al. 1999; Dessus et al.

1999; Nagarajan 1999; ABARE 2000; Anderson et al. 2000; Hertel et al. 2000; Brown et al. 2001; Diao et al. 2001; Francois 2001; Scollay and Gilbert 2001; vanMeijl and vanTongeren 2001; World_Bank 2001). The authors summarize key findings of twelve CGE assessments in a table (their table 11) that is (slightly modified) reproduced in figure 1 below. From a modeler's point of view the large variance between results may not be surprising, or, as UNCTAD (2003) states related to their table 11*: "First of all, the global welfare results are quite similar across models and studies".* However, the demand side for these results may not easily find an intuition for the fact that presented welfare changes for a complete (100%) liberalization of all agricultural markets world wide are roughly half as big as world wide gains from a 50 % liberalization of agricultural markets (compare figure 1). In addition, simulated welfare changes for a comprehensive liberalization in all markets and sectors of the world economy range from 78 billion US$ to 1857 billion US$ and make it hard to derive at least a rough average (compare figure 1).The authors of UNCTAD (2003) address this issue by providing explanations for the behavior of various assumptions incorporated into the CGE models under review; some of these features have been included into figure 1 in parenthesis behind the author's name. The authors of UNCTAD also mention that *"… a consensus is emerging among modelers that, owing to the robustness problems […], results from CGE analysis should be interpreted more in a qualitative than in a quantitative sense, and that putting too much emphasis on specific numbers and figures should be avoided"* (UNCTAD 2003), citing (Francois 2000).

It is very important to communicate the limitations of simulation experiments to the demand side of simulation results; however, from a methodological point of view it is not efficient to focus on the qualitative interpretation of experiments that have been designed to be quantitative. Especially within the political environment of trade negotiations, it will be crucial not only to identify winners and losers from liberalization but also to quantify the regional distribution of burdens and benefits from trade policy changes in order to make the negotiation of potential compensations possible, transparent and credible. The growing demand for quantitative results will make it hard for economists to keep simulation results within the interpretational framework that they consider appropriate. It might rather happen that:*„ …the results of numerical analysis can take on a life of their own when released into a policy environment. Very rough guesses can be marketed as precise estimates, packaged with interpretations […] unintended by the original analysts. In such a situation, the limitations of the original analysis are easily (and sometimes deliberately) forgotten. For this reason, it is very important that the nature of the analytic exercise be kept in mind, and be transmitted […] to the relevant policy audience. Transparency is critical."* (Francois and Reinert 1997, p.20)

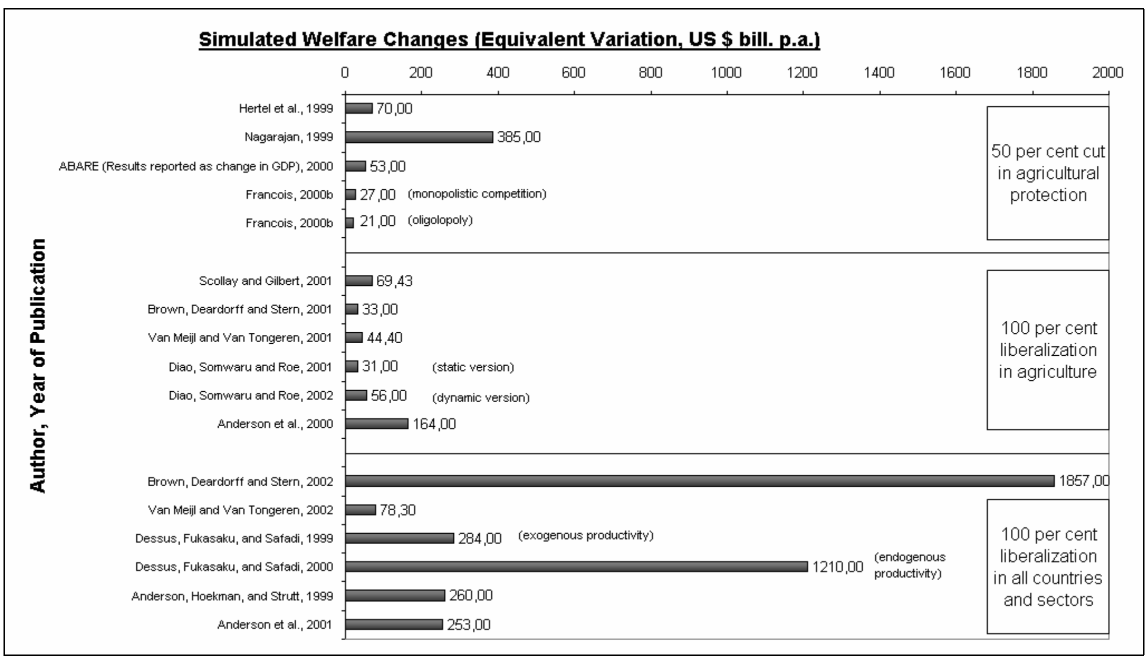

Figure 1: Results of Doha Round CGE simulations. Source: own depiction based on UNCTAD (2003).

Obviously, simulation results should not be interpreted outside the context of their experimental settings. Quantitative results that cannot be compared to each other in a reasonable way may not contribute to liberalization discussions but rather raise concerns about further liberalization efforts. The “black box” character and the unexplained variance as shown in figure 1 may strengthen skepticism about economists’ motivation behind their experiments and tend to worsen the asymmetric information gap.

Figure 2 narrows the observations presented in figure 1 down to simulated effects from various scenarios of agricultural trade liberalization as they might affect the European Union. Similar to figure 1, the variance between simulation results (EV) cannot be explained by the shock size, even if region and sector are held constant. All observations are measured as equivalent variation (EV), except the results presented by ABARE (2000), which presents welfare effects as changes of GNP per capita relative to the base level. Although the results are numerically in range with the reported EV effects from all other publications within the sample, it is important to note that these two measures are not directly comparable within the framework of CGE models. Thus, presenting both measures next to each other may lead to misleading conclusions.

To some extend, the authors of UNCTAD (2003) provide explanations for the variance between the CGE assessments presented in Figure 1 and Figure 2. Since most studies present simulated welfare changes as EV (or CV, see (Varian 1992)), we use this measure in order to compare the nature of the various CGE assessments specifically reviewed by UNCTAD (2003). In order to ensure strict comparability, the study by ABARE (2000) had to be excluded, since no information on EV is provided. Additionally, the study by World Bank (2001) has been excluded since no specific details were provided on the nature of their simulation results. Instead, the results presented by the authors of UNCTAD (2003) have been included into the sample, although the authors do not present these simulation results along with their qualitative comparison of other simulations. The publication by Francois (Francois 2001) is a monograph identical to the publication cited by UNCTAD (2003) except, a different title applies (Francois 2004).

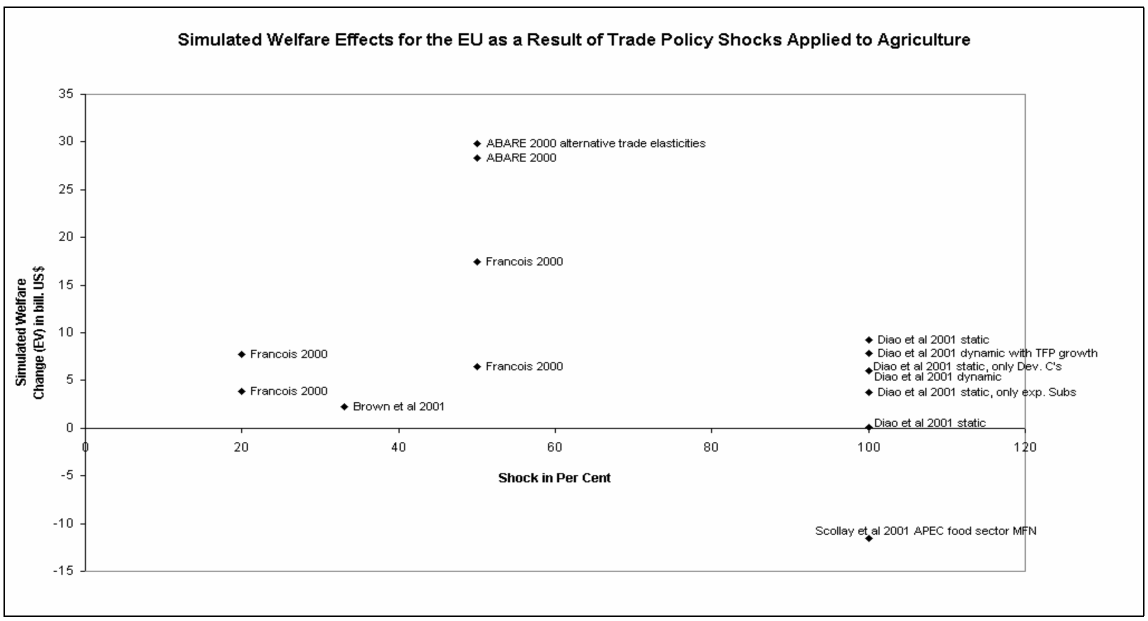

Figure 2: Simulation Results for the EU, policy changes in agriculture only. Source: Own Depiction.

Out of the simulation results presented within these publications, we compile a dataset that contains 11 simulation studies (s), including 60 different experiments (x) and in total 719 observations (n) on welfare changes (EV) at the regional level (r). No experiment has been excluded since we assume that the authors present their simulation results in order to address a specific policy issue, in this case trade issues under the Doha round, rather than methodological comparisons. Thus, all published experiments are considered to be potentially relevant for the “Doha round”. The n = 719 observations stem from 76 regions (r), out of which we have rated 17 as identical or very similar. For example, it has been assumed that the group of “high income countries” presented by Anderson et al. (1999) is equal to the group of OECD countries. We have also assumed that any group of “developed countries” can be treated equally to the group of OECD countries, unless the authors have stated something different. These assumptions have been necessary in order to approximate the average GNI per capita level for the various country groupings. This is for the following reason:

Any simulated change (increase) of the dependent variable “welfare” does not necessarily imply a large absolute gain in welfare; for instance, an estimated EV of 1 billion US$ would shift the utility based indifference curve for Bangladesh comparatively far out to the right, meanwhile the same simulated welfare change is of only minor impact on utility in the USA or Japan

We have used GNI per capita data for the year 2003 in order to identify the relative magnitude of the predicted welfare changes according to an economy’s state of development. However, due to the complexity of this task and missing information on the precise regional aggregation undertaken by the authors, we can only approximate the arithmetic average of GNI per capita in (r). The base year for most of the databases in use is either 1992, 1995 or 1997, and at the same time most simulations predict their database to the year 2005. Thus, the year 2003 has been chosen as an approximation to average income level in (r).

All reported “total effects” or “world in total” figures except the ones for van Meijl and van Tongeren (vanMeijl and vanTongeren 2001) have been removed from the dataset since these observations present weighted sums of all other regional observations. In case of van Meijl and van Tongeren, no absolute welfare changes have been reported at the regional level. We have included the simulation results for “world” from this publication as a single observation. An inclusion of these weighted sums for total world welfare effects would lead to an overrepresentation of countries with large impact on world welfare.

In many CGE simulations, the ROW (“rest of the world”) aggregations within CGE simulation bear little economic message, or, in different words, whatever happens in the ROW is a broad aggregation over countries that are very often neither comparable to each other by income- nor by protection levels and thus are hardly ever within the focus of any trade policy analysis. Nevertheless, we have included the ROW aggregations into our dataset in order to avoid a potential bias.

Within the n = 719 observations, the three largest welfare losses in absolute terms occur in an aggregated region r = “all developing countries”, for an experiment that analyzes a removal of distortions in OECD countries for the manufacturing sector. The second largest losses are simulated for Europe in case of APEC internal food liberalization, the third largest losses exhibit the USA under an simulated 33% reduction of agricultural tariffs worldwide. Contrary to these results, the largest welfare gain within our sample is shown for all OECD countries under complete liberalization of all markets in all sectors worldwide.

In order to group related information on the various CGE models into category variables, the framework established by van Tongeren and van Meijl (2001) has been used. The authors present a tree diagram that groups contemporary CGE models according to important characteristics. For example, the inclusion of the Armington specification distinguishes models from those with imperfect competition. Further, the nature of a model is static, recursively dynamic or dynamic. For example, according to van Tongeren et al. (2001) , we have grouped recursive models as static by nature and have included a recursive dummy variable, meanwhile other models have been grouped as purely dynamic.

The information on the experimental settings were grouped into mostly binary variables. In case of missing information, the authors of simulation studies have not been contacted. Instead, either the variable has been categorized on a broader level in order to include information from all studies or the variable has been dropped from the dataset. Table 1 at the end of this paper provides an overview on the characteristics of the dataset that we have compiled from the 11 CGE simulations in focus. Generally, we have assumed that no documentation of any specific issue implies that the authors do not make assumptions on this specific modelling characteristic that influences reported welfare changes significantly or, otherwise, they do not deviate form specifications of the standard GTAP model. A complete dataset with all observed variables is available on request.

4. Results

The OLS linear model is both a flexible and robust approach and at the same time comparatively easy to interpret. The Breusch-Pagan test has rejected the H_0 hypothesis of homoskedasticity within our dataset at the 5% level of probability. For this reason, we have estimated a heteroskedasticity consistent variance- covariance matrix and have applied robust standard errors to the estimated coefficients shown in table 2.

The model has been designed according to the criteria established by authors of qualitative model comparisons (Francois 2000; van Tongeren et al. 2001; Hertel 2002; UNCTAD 2003). The variables from our dataset follow the classification tree established by (van Tongeren et al. 2001:169) "downwards" from a general to a specific level. The variables have stepwise been included into the model and tested according to their significance and contribution to the overall fit.

In order to identify the potential impact of a simulated welfare change on its economy, data on gross national income per capita in the year 2003 ("gni03") have been included into the regression. The estimated coefficient is positive at an estimated 0.00047 billion US$ gain in EV per additional US$ gross national income per capita. This result should not lead to the conclusion that "rich" countries gain more from liberalization according to the models within the sample; since the dependent variable is measured in absolute terms, it would be surprising to find a negative sign for this coefficient since large economies with high levels of protection will ceteris paribus gain most in terms of allocative efficiency and, sometimes, terms of trade (consider, for instance, a simulated CAP reform for European Union).

The interpretation of the "shock" variable is strait forward: a 100% shock is the maximum degree of liberalization that is possible. No matter if a region (r) is "rich" or "poor" in terms of income per capita, the assessed models tell us that an increase in liberalization will increase the size of the expected welfare gain. The coefficient indicates that 1% simulated cut of protection in trade (mostly, only tariffs are considered in the sample simulations) is correlated with an increase in a region's EV by 0.158 billion US$. Most of the liberalization scenarios in our sample take place according to the pattern "within all countries and for all countries". However, some experiments employ scenarios that are closer to reality; they shock the model only for free trade agreements (FTA) between certain regions. We have grouped this information in dummy variables but found none of these dummies to be significant within our regression. Obviously the effects of more fundamental variables have stronger explanatory power on the dependent variable than the regional setting of the shock. The sectoral dummies indicate whether a policy change occurs in agriculture, manufacturing or all sectors at the same time; none of these variables is significant within our model.

The "dynamic" dummy is the first variable out of the classification tree by van Tongeren, van Meijl et al. (van Tongeren et al. 2001) that proved to be significant within the context of variables observed at a more detailed level. Dynamic simulations show an estimated increase on simulated welfare changes within our sample: Ceteris paribus, a CGE will shift the intercept by 217 billion US$ EV for any region (r) if the model is dynamic (compared to static or recursive models). Slightly larger appears to be the impact of increasing returns to scale: if employed in the model, this assumption shifts the intercept of observed welfare changes in a region up by 261 billion US$ relative to models that employ constant returns to scale. These estimates are in line with the qualitative knowledge documented on the effects of these assumptions (Francois and Roland-Holst 1997).

The University Dummy "uni" is the most significant variable out of a group of dummies that trace the origin of a study (compare table 1). Clearly, our sample of 11 studies does not allow for any general conclusions on the broad magnitude of simulation results according to their institutional origin. Most likely, this points towards an unobserved feature that these four studies have in common (compare table 1). However, these publications show simulated welfare changes on average 66 billion US$ higher than for simulations from studies that have been conducted not by a purely academic institution.

The "databaseyear" variable ranges from 1992 to 1997. All publications employ a version of the GTAP database. Older versions of GTAP data account for larger welfare gains due to the fact that tariff reductions under the Uruguay round have not been phased in (UNCTAD 2003). In addition, the year 1995 as the base year for GTAP 4 dataset has been market with high world prices in agriculture and thus incorporates comparatively low levels of protection (Hertel et al. 2000). According to this, it is obvious that a projection to a year more distant into the future, such as 2010, is also negatively correlated with welfare changes since the experiments usually assume continued liberalization effort that brings protection levels linearly down on a yearly base. In other words, an experiment that employs "young" GTAP data or data with high world prices and projects policy shocks far into the future will ceteris paribus produce comparatively smaller simulated welfare gains.

The "ChinaWTO" dummy is 1 if the simulation experiment assumes that China and Taiwan join the WTO before 2005 and gain access to OECD markets just as all other developing countries already being members of the WTO. Intuitively, the estimated coefficient for this dummy appears to be high at 210 billion US$ for an average region. Out of eleven studies, four within our dataset employ this assumption (compare table 1). However, we do not have information from all publications within our sample on how they treat this trade policy assumption in detail. (Diao et al. 2001) explicitly assume that no further liberalization will take place in China. The authors state that *"…If China liberalizes agriculture, the level of world agricultural prices would rise by 12.2 percent, an increase of about 0.6 percent over [our] non-China predicted values."* (Diao et al. 2001:3). The significance of this variable points to the fact that assumptions about the political environment can be at least as influential on simulation results as assumptions about behavioural equations or exogenous parameters. This is especially crucial since none of the publications within our survey explicitly analyses China's WTO accession. Rather, all studies focus on prospective Doha- round scenarios; the accession of China is only mentioned as a side aspect.

Table 2: Linear model of simulated welfare changes in billion US$ EV observed at the regional level.

```
Linear Regression Model

         n Model L.R.       d.f.         R2      Sigma
       719       1226         14     0.8182      29.42

Residuals:
     Min       1Q   Median       3Q      Max
-182.595   -9.264   -2.616    4.620  272.879

Coefficients:
                        Value    Std. Error      t    Pr(>|t|)
Intercept         118769.71253   27749.13868   4.280    0.000
GNI03                  0.00047       0.00009   5.225    0.000
shock                  0.15890       0.06585   2.413    0.016
dynamic              216.99351      48.17970   4.504    0.000
uni                   66.30785      15.87651   4.176    0.000
returnstoscale       261.15975      53.64107   4.869    0.000
databaseyear         -41.14074       9.87921  -4.164    0.000
projectionyear       -18.38658       4.52888  -4.060    0.000
chinawtodummy        210.61666      41.64764   5.057    0.000
newgrowth             33.81241       9.44724   3.579    0.000
newtrade            -254.58637      52.07584  -4.889    0.000
armingtondouble     -200.62857      37.03369  -5.417    0.000
regions                4.42841       1.61125   2.748    0.006
sectors                4.91639       0.75963   6.472    0.000
welfab290            427.96598      42.87622   9.981    0.000

Residual standard error: 29.42 on 704 degrees of freedom
Adjusted R-Squared: 0.8146
```

The “newgrowth” dummy equals 1 if a model incorporates features of the New Growth theory. These features for example can either be a model closure with endogenous technical change or exogenously specified technical spillover effects, or any other feature related to the principles of the New Growth Theory. Due to limited information provided along with the documentation of the eleven studies in focus, the nature of this variable is broad and does not distinguish between entire modeling concepts or rather small features. Nevertheless, the coefficient indicates an intercept shift of 33.8 billion US$ EV within the average region (r) if a model exhibits new growth features.

Very similar, our “newtrade” dummy groups features that can broadly be summarized under the framework of the New Trade Theory. Although increasing returns to scale are an important assumption for markets in which firms have oligopolistic or monopolistically competitive market power, the variable must be distinguished from the “returnstoscale” variable in our model. The latter addresses general deviations from the standard CGE assumptions (van Tongeren et al. 2001), meanwhile the “newtrade” dummy rather incorporates simulation experiments with build-in features regarding key markets in focus of the simulation experiment. For example, the model used by Francois (2000) initially shows constant returns to scale and perfect competition for some sectors although for other sectors oligopolistic markets and monopolistic competition are modeled through extensions to the base model (Francois 2000). These extensions are limited to certain experiments. The coefficient is negatively correlated with simulated welfare changes in the average region. The downward shift at -254 billion US$ most likely results from the fact that (Brown et al. 2001) and (Francois 2000) present several scenarios with moderate policy shocks where negative welfare changes occur for some regions. Due to the fact that these authors use datasets disintegrated into comparatively many regions and we have not weighted the observations according to GNP or the share of total observations within our dataset, these scenarios influence regression results strongly. However, the negative impact of new trade features in general does not necessarily contradict findings of other studies. For example (Elbehri and Hertel 2004) employ a model with oligopolistic markets and increasing returns to scale for their analysis of a Morocco- EU FTA. The authors find deteriorating terms of trade to outweigh minor allocative efficiency gains for Morocco. Francois (2000) also discusses the importance of terms of trade effects for the outcome of simulation results on the regional level. On the other hand, ceteris paribus, the pro competitive gains from new trade features should be high (Francois and Roland-Holst 1997).

The “armingtondouble” dummy is one if a study states an increase or doubling of their Armington elasticities without additional econometric analysis concerning this parameter. Increasing the Armington elasticities has evolved as a tradition among users of the GTAP modeling framework since the assessment by (Gehlhar 1997) has found ex post that increased Armington elasticities better explain trade patterns that actually have been observed in reality. However, as (Hertel et al. 2000) point out, there is little theoretical justification for doing so. As in the previous case, the coefficient for this variable is negative at 200 billion US$.

Estimates for the number of aggregated regions and aggregated sectors (“regions”, “sectors”) within the dataset are strait forward in their interpretation. Theoretically, for each region added to the aggregation scheme and for each sector, a model in our sample will show 4.4 billion and 4.9 billion US$ of welfare gain within the average region (r), respectively. Increasing the degree of regional and sectoral aggregation may even out protection levels and lowers the absolute size of terms of trade effects.

“Welfab290” is an outlier dummy. We found that very few observations are responsible for a major share of variance within the dataset (compare the quartile distribution of the residuals in table 2). We have grouped 11 observations into this variable that all show simulated welfare changes above 290 billion US$. An inclusion of even more outliers into this variable has not improved the overall fit of the model. The outlier dummy addresses a small sub sample of the dataset with average welfare gains at 432 billion US$, all other variables assumed to be zero. To some extend, the large absolute values estimated for some of the dummy variable coefficients in our model (as discussed above) are due to these outliers. An alternative to the specification of this dummy would be the selection of two (or more) different variable subsets. However, in order to provide a comprehensive assessment of the literature review provided by UNCTAD (2003), we have chosen the specification of outlier dummies in order to avoid further questions concerning the selection of an optimal sub sample.

As for any other variable in the model, we cannot draw conclusions on the impact of new trade features on modelling results in general. We rather have to limit our interpretation to causal relationships within our dataset.

Instead, the interaction of all modeling assumptions will determine the outcome of the simulation. The estimated coefficients presented in table 2 and 3 do not contradict the simulation results for "total welfare" or "whole world" as presented in figure 1. As explained in the previous section, these aggregated observations have not been included into our dataset.

5. Discussion

Three important results emerge from our analysis: first, the explanatory variables within our econometric model explain a major share of the variance within the sample. Secondly, these variables address important modeling features discussed in quantitative reviews and the estimated signs are in range with those qualitative predictions. Thus, the econometric representation is a feasible way to summarize simulated regional welfare effects in relation to important modeling assumptions. As a third important result, the analysis has highlighted the policy dummy for a WTO accession of China and Taiwan before 2005 as a significant factor influencing welfare changes within our sample. This is especially important since this policy assumption is not within the focus of simulated scenarios and may easily be ignored or overseen by the demand side of simulation results.

These three findings underline the fact that simulation results of CGE based policy assessments within our sample cannot be interpreted in a reasonable way if only the size of the policy shock is considered. Rather, many different factors concerning the experimental setting, exogenous parameters and the assumed policy environment have to be taken into account and prove to be at least equally influential on the simulation results as the size and nature of the policy shock. All these features have to be discussed along with simulation results in order to avoid "black box" frustration among those who have to focus on the magnitude of simulation results rather than on the nature of simulation experiments.

For economists working with CGE models, our analysis initially presents little new insight; however, our approach would potentially enable better comparisons of various CGE models relative to each other. Specific model components could be tested according to their relative impact on each other within the same model or across different models. Due to limited information on the details of each single assessment, our variables express rather broad directions since most variables take only binary values. Clearly, one may think of other variables that could be tested within an econometric model of CGE- based results, for example the absolute value of different Armington elasticities. While sensitivity analysis is a tool to assess the functioning of parameters within a specific CGE model (Hertel et al. 2004), our econometric approach may enable the construction of confidence intervals across models. The assumed linear relationship does not capture the effects that different modelling features unfold at various hierarchical levels of the dataset. For this reason, a larger dataset is required that provides more observations on the various functional levels of a CGE model. With such a dataset, a more sophisticated econometric model that accounts for the nested structure of information within the dataset can be applied and will most likely present further insight into the role and interaction of important modelling assumptions within a CGE framework.

The most appropriate econometric approach to capture the "nested" structure of a dataset would be the application of a multi- level model (Goldstein 1995). Multi- level models employ the following principle:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_{ij} \tag{3}$$

$$\beta_{0j} = \beta_0 + \mu_j \tag{4}$$

with
i = publication i for I publications, i= 1,…,n ; within the dataset the first level.
j = modelling framework (e.g. GTAP, BDS), j= 1,…,n describes the second hierarchical level.
β_n = coefficients,
u, μ = error terms, cov $(u_j , u_{j'})$ = cov $(\mu_{ij} , \mu_{i'j})$ = cov $(u_{j'} , \mu_{ij}) = 0$
y = simulated welfare changes in region (r); the (r) index has been omitted for simplicity.
x = explanatory variable, e.g. policy shock.

Equation (4) introduces an additional constant; β_0 describes the component common across all observations and μ_j is the disturbance that is introduced by model j. Plugging (4) into (3) yields equation (5):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \mu_j + u_{ij} \tag{5}$$

Thus, a multi- level model splits the disturbance term into components that contain information about the various observational levels: μ represents variance within the dependent variable that is due to the modelling framework, meanwhile u accounts for variance within the dependent variable itself. If $Var(\mu_j) > Var(u_{ij})$ then the variance within y will mostly be due to the modelling framework; if $Var(u_{ij}) > Var(\mu_j)$ it will be rather due to other factors. Multi- level models require observations for all observational levels. Especially the effects of various degrees of imperfect competition and the interaction with scale economies could potentially be explored with multi- level methods if more observations on the publication level would allow for comparisons.

On the other hand, our approach might be even more useful as a "rule of thumb" comparison for several CGE experiments in focus of a specific trade policy issue, such as the "Doha" round in our application. Our dataset is based on 12 publications that have been selected by the authors of UNCTAD (2003) in order to support a political discussion ("Back to Basics") rather than a methodological comparison of CGE features. However, the results as presented in figure 1 will most likely tend to lower overall confidence in the results of CGE- based assessments. If the results of CGE- based policy assessments could gain credibility among a broader audience, it is very likely that CGE models would contribute even more to the ongoing trade negotiations. In other words: Improved efficiency for information based on CGE models could provide missing information on the nature of trade policy shocks and thus might potentially speed up trade negotiations itself. However, the task to improve CGE models as scientific tools to conduct economic experiments will remain at the core of CGE- model related research.

The suggestion by some authors to interpret simulated quantities in a qualitative way cannot efficiently solve the asymmetric information problem for trade policy related analysis. Instead, our econometric approach to proxy the experimental settings could provide the demand side of trade policy related information with a quantitative "roadmap" to the results of different CGE assessments.

This paper has outlined a promising method to establish this "roadmap" and shed light on the variance between simulation results (figure 1). A replication of our analysis does not require any insight knowledge on each model's detailed structure, since only information has been used that is documented along with each publication included into the sample. However, the econometric results can only be regarded as a first step and provide a starting point for further research.

Table 1: Publications included into the dataset; a variable takes the value 1 if the specific feature is present.

Publication	Model's "Family"	Organization	dynamic	Returns to scale, 1 = increasing	Base Year	Projection Year…	"China WTO"	New Growth Theory	New Trade Theory	Armington doubled ?	Reg's	Sec's
ABARE 2000	GTEM	Nat.Org.	1	0	1995	2010	0	1	0	1	17	24
Anderson et al 1999	GTAP	Int.Org and Univ	0	0	1992	2005	1	0	0	1	12	5
Anderson et al 2000	GTAP	Int.Org and Univ	0	0	1995	2005	1	0	0	0	2	4
Brown et al 2001	"Michig. Trade"	Univers.	0	1	1995	2005	0	0	1	0	21	18
Dessus et al 1999	"OECD"	Int.Org and Univ	1	0	1995	2010	0	0	0	0	16	4
Diao et al 2001	"WTO"	Int.Org and Univ	0	0	1997	1997	0	1	0	0	12	9
Francois 2000	GTAP	Univers.	1	0	1995	2000	0	0	1	0	9	19
Hertel et al 1999	GTAP	Univers.	0	0	1995	2005	1	0	0	1	19	22
Nagarajan et al 1999	GTAP	Nat. Org.	0	0	1995	1995	1	0	1	0	9	9
UNCTAD 2003	GTAP	Int. Org.	0	0	1997	1995	0	0	0	0	12	6
Scollay et al 2001	MRT	Int.Org and Univ	0	1	1995	2005	0	1	0	1	15	15
VanMeijl et al 2001	GTAP	Univers.	0	0	1997	1997	0	0	0	0	15	11
World Bank 2001	dropped from sample due to lack of comparable information											

References

ABARE (2000). The Impact of Agricultural Trade Liberalisation on Developing Countries. Canberra, Australian Bureau of Agricultural and Resource Economics. Authors: Fran Freeman, Jane Melanie, Ivan Roberts, David Vanzetti, Apelu Tielu, Benjamin Beutre.

Aisbett, E. (2004). Why are the Critics so Convinced that Globalization is Bad for the Poor? Paper presented at the NBER Globalization and Poverty Conference, September 10-12.

Akerlof, G. (1970). The market for "lemons": quality uncertainty and the market mechanism. The quarterly journal of economics, MIT Press 84(3): 488-500.

Anderson, K., B. Hoekman and W. Martin (2000). Potential Gains from trade Reform in the New Millenium. Paper presented at the third Annual conference on Global Economic Analysis, Monash University, June 27- 30.

Anderson, K., B. Hoekmann and A. Strutt (1999). Agriculture and the WTO: The Next Steps. Paper prepared for the Second Annual Conference on Global Economic Analysis, Avernaes Conference Centre, Helnaes, Denmark, June 20-22.

Bertelsmeier, M., W. Kleinhanss and F. Offermann (2003). Aufbau und Anwendung des FAL-Modellverbunds für die Politikberatung. Agrarwirtschaft 4(8): 175-184.

Brown, D. K., A. V. Deardorff and R. M. Stern (2001). CGE Modeling and Analysis of Multilateral and Regional Negotiating Options. Medford, Tufts University: 50.

Dawkins, C. (1997). Extended Sensitivity Analysis for Applied General Equilibrium Models. Warwick Economic Research Papers No. 491. Warwick.

Dessus, S., K. Fukasaku and R. Safadi (1999). Multilateral Tariff Liberalisation and the Developing Countries. Paris, OECD Development Centra: 36.

Diao, X., A. Somwaru and T. Roe (2001). A Global Analysis of Agricultural Reform in WTO Member Countries. Washington D.C., Economic Research Service/USDA: 40.

Elbehri, A. and T. Hertel (2004). A comparative Analysis of the EU- Morocco FTA vs. Multilateral Liberalization. GTAP Working Paper No. 31.

Francois, J. (2000). Assessing the Results of general Equilibrium Studies of Multilateral Trade Negotiations. New York and Geneva, UNCTAD Study Series No. 3.

Francois, J. (2001). The next WTO Round: North-South stakes in new market access negotiations. Adelaide, Rotterdam, Amsterdam, Centre for International Economic Studies. University of Adelaide, Tinbergen Institute: 77.

Francois, J. (2004). Email correspondence, December 2004.

Francois, J. and K. A. Reinert, (eds.) (1997). Applied Methods for Trade Policy Analysis- A Handbook, Cambridge University Press.

Francois, J. and D. Roland-Holst (1997). Scale Economies and Imperfect Competition. In J. Francois and A. Reinert, (eds.) Applied Methods for Trade Policy Analysis. Cambridge University Press.

Gehlhar, M. (1997). Historical analysis of growth and trade patterns in the Pacific Rim: an evaluation of the GTAP framework. In T. W. Hertel, (ed.) Gloabl Trade Analysis. Cambridge, Cambridge University Press: Chapter 12, p. 349 - 363.

Goldstein, H. (1995). Multilevel Statistical Models. Arnold, London- Sydney- Auckland.

Heisenberg, W. (2000). Physik und Philosophie, Hirzel.

Hertel, T., D. Hummles, M. Ivanic and R. Keeney (2004). How Confident Can We Be in CGE-Based Assessments of Free trade Agreements? GTAP Working Paper No. 26.

Hertel, T. W. (1999). Future Directions in Global Trade Analysis. Department of Agricultural Economics, Purdue University.

Hertel, T. W. (2002). Applied General Equilibrium Analysis of Agricultural And Resource Policies. In B. Gardener and G. Rausser, (eds.) Handbook of Agricultural Economics, Elsevier Science B.V. 2.

Hertel, T. W., K. Anderson, J. F. Francois and W. Martin (2000). Agriculture and Non-Agriculture Liberalization in the Millennium Round. Adelaide, Centre for international economic studies: 33.

Kehoe, T. J. (2002). An Evaluation of the Performance of Applied General Equilibrium Models of the Impact of NAFTA. Paper Presented at the 5th Conference on Global Economic Analysis www.econ.umn.edu/~tkehoe.
Nagarajan, N. (1999). The millennium round: An economic appraisal. Bruxelles, Luxembourg, CECA –CEE - CEEA: 43.
Nölle, F. and M. Banse (2001). Wie gut lassen sich Agrarpolitische Reformen durch partielle Modelle abbilden? Eine ex-post Analyse der MacSharry Reform am Beispiel von ESIM.In M. Brockmeier, F. Isermeyer and S. von Cramon-Taubadel, (eds.) Liberalisierung des Weltagrarhandels- Strategien und Konsequenzen. Münster-Hiltrup, Landwirtschaftsverlag.
Scollay, R. and J. Gilbert (2000). Measuring the Gains from APEC Trade Liberalization: An Overview of CGE Assessments. The World Economy 23(2): 175-197.
Scollay, R. and J. Gilbert (2001). An integrated approach to agricultural trade and development issues: Exploring the welfare and distribution issues. New York and Geneva, APEC Study Centre and Economics Department, University of Auckland New Zealand, Dep. of Agr. Economics, Washington State University: 43.
UNCTAD (2003). Back to Basics: Market Issues in the Doha Agenda. New York and Geneva, Authors: S. Laird, L. Cernat, A. Turrini.
van Tongeren, F., H. van Meijl and Y. Surry (2001). Global Models Applied to Agricultural and Trade Policies: A Review and Assessment. Agricultural Economics 26: 149-172.
vanMeijl, H. and F. vanTongeren (2001). Mulilateral trade liberalisation and developing countries: A North-South perspective on agriculture and processing sectors. West Lafayette, Purdue University: 34.
Varian, H. R. (1992). Microeconomic Analysis, Third Edition. New York/London, Norton.
World Bank (2001). Global Economic Prospects and the Developing Countries. Washington D.C., World Bank.
WTO (2004). The Future of the WTO: Adressing institutional challenges in the new millenium. Geneva, World Trade Organization.