Meta-Analysis of General and Partial Equilibrium Simulations of Doha Round Outcomes

Sebastian Hess and Stephan von Cramon-Taubadel[1]

"For were it not better for a man in fair room …
branching candlestick of lights, than to go about with a
small watch- candle into every corner?"
Francis Bacon (1605)

Abstract:

Quantification of welfare changes due to trade liberalisation play a crucial role for political decision making. However, meaningful comparisons of simulation results from different sources are difficult. Often significant differences in simulated gains from liberalisation do not serve to increase confidence in quantitative assessments based on trade models. We employ a meta-analysis of applied trade simulations under the WTO Doha Round to identify model characteristics that influence the magnitude of simulation results, and to estimate the magnitude of this influence. Findings from our simple econometric model are plausible and show that each simulation experiment represents a complex interaction of experimental settings that may not easily be understood by and communicated to non-experts. Meta-analysis proves to be a useful tool for empirically assessing this complexity.

Keywords: Meta-analysis, CGE, Partial Equilibrium, Trade Liberalization
JEL: C00, C23, C68, F10

1. Introduction

Ongoing debates about the pros and cons of further agricultural trade liberalization often hinge on empirical estimates of the gains and losses that would accrue to specific interest groups, countries and regions. Applied trade models provide such estimates and have become an important part of the political decision making process (Devarajan and Robinson 2002).

[1] Department of Agricultural Economics and Rural Development, University of Göttingen. The authors are grateful to Yves Surry, Bernhard Brümmer, Martin Banse and Frank van Tongeren for many useful suggestions, and to Tinoush Jamali and Inken Köhler for valuable research assistance. This project has been supported by the German Research Council (DFG), and Sebastian Hess has partly been supported by the Friedrich-Naumann-Foundation.

However, applied trade models are frequently criticised as having weak empirical foundations (Alston, Carter, Green and Pick 1990; Alston, Norton and Pardey 1995; McKitrick 1998; Anderson and Wincoop 2001) and as being insufficiently transparent (Ackerman 2005; Piermartini and Teh 2005). In addition, different models often produce trade simulation results that "… differ quite widely even across similar experiments" (Charlton and Stiglitz 2005), and convincing explanations for these differences are, due to the complexity of many models, often difficult to provide. These problems complicate an already difficult debate on agricultural trade liberalisation, and are water on the mills of those who cast doubt on the benefits of liberalisation and the ability of economists to provide objective measures of these benefits.

In this paper we present a meta-analysis of partial and general equilibrium results on the Doha Development Round (DDR) of WTO negotiations. The aim of this analysis is to identify model characteristics (e.g. partial vs. general equilibrium, level of disaggregation) and other factors (e.g. the database employed) that influence simulation results in a systematic manner, and to derive quantitative estimates of these influences. To the degree that this work-in-progress is eventually successful, we hope that it will contribute to a more transparent debate on the pros and cons of liberalisation and to our ability as a profession to inform this debate.

This paper is organized as follows. In Section 2 we briefly review methods for comparing and evaluating quantitative simulation results, and we discuss the strengths and weaknesses of meta-analysis as a tool measuring the impact of variables that influence trade liberalisation simulation results. Section 3 contains information on the collection and processing of the dataset employed in our meta-analysis. Section 4 presents preliminary results which are discussed in Section 5 with regard to their implications for policy making and for further research.

2. Approaches for comparing and evaluating simulation results

2.1 Past approaches

Two main approaches have been employed to compare simulation results from trade models. The first approach is sensitivity analysis. Sensitivity analysis measures the sensitivity of simulation results to variations in selected parameters and model specifications (Hertel, Hummels, Ivanic and Keeney 2003). A potential shortcoming of sensitivity analysis is that modellers will presumably have few incentives to report findings that indicate that their simulation results are highly sensitive to certain model characteristics (Stanley 2005); instead, there will be a natural inclination to demonstrate that simulation results are 'robust'. Sensitivity analysis can cast light on the characteristics and performance of an individual model, but it cannot be readily applied to compare results from different models and experiments. Hence, sensitivity analysis is usually confined to a specific modelling framework. An example is provided by Gohin and Moschini (2005), who use the GTAP model to compare general equilibrium (GE) with partial equilibrium (PE) closures.

The second approach used to compare the results of trade model simulations is the qualitative review (Robinson 1989; Scollay and Gilbert 2000; Laird, Cernat and Turrini 2003; Charlton et al. 2005; Piermartini et al. 2005). In such reviews, tables and graphs are often used to compare results from different models classified according to selected model characteristics ('dynamic model', 'increasing returns to scale', etc). A shortcoming of this approach is that such essentially bivariate comparisons can be misleading if variations in other model characteristics, in liberalisation experiments, and in databases between studies are not controlled for (Harrison, Rutherford and Tarr 1997).

2.2 Meta-analysis

We employ meta-analysis to compare quantitative results from applied trade models and thus generate insight into how model characteristics influence these results. Meta analysis is a comparatively recent inductive empirical method that seeks to find similarities and explain differences between scientific findings on similar research questions across publications (Stanley and Jarrel 1989; Chalmers and Altman 1995). Meta-analysis has three objectives:

- Combining evidence: This was the primary goal when the method was established in the fields of medicine and psychology.
- 'Separating wheat from chaff': In some meta-analyses, groups of experts devise weighting schemes for the scientific quality of publications within the meta-sample. If the studies in such a sample report widely differing results, it is important to identify those that deserve a higher weighting because they can be considered more reliable.
- Evaluating methods (Stanley 2001; Florax, de Groot and de Mooij 2002): This approach has evolved especially in economics and related disciplines in which reproducible measurements are often hard to obtain and quantitative results are known to depend heavily on the methods that have been applied. Meta-analysis can quantify the share of variance within a given set of estimates that is due to different methodologies and *a priori* assumptions.

Each meta-analysis also faces at least three major methodological threats that may cause the method to fail, leading to biased results:

- 'Comparing apples and oranges' (Eysenck 1995): This occurs when the studies being analysed utilize approaches that are so fundamentally different that no common evidence can be identified.
- 'Junk in junk out' (Wachter 1988): If standards in a certain research area are, for whatever reason, poor in the first place, a meta-analysis cannot be expected to yield evidence that is of any higher value.

- Publication bias (Stanley 2005): Meta-analysis hinges on the selection of an appropriate literature sample. The results of a meta-analysis will be biased if strands of the literature are underrepresented in the sample. This can be due to inadequate search strategies, but also due to a bias against publishing statistically insignificant results. It might also be that an important sub-set of the available evidence is published in obscure, non-English journals. Finally, some relevant studies might be confidential and thus unavailable for meta-analysis.

If these pitfalls can be avoided, meta-analysis has the potential to improve on sensitivity analysis and qualitative reviews by permitting comparisons of the results of trade model simulations that control for simultaneous variations in model characteristics, experimental settings, data bases and other study characteristics. The meta-analysis that we propose to carry out is based on the following general model:

$$I_i = f(MC_i, LE_i, DB_i, SC_i, u_i), \qquad (1)$$

where I is the simulated impact of a trade liberalisation experiment, MC is a vector of model characteristics (such as partial or general equilibrium), LE describes the liberalisation experiment (the magnitude of the simulated tariff reductions), DB is the database underlying the simulation (e.g. GTAP 4 or 5), SC is a vector of study characteristics (for example affiliation of the authors, whether the study has been published), u is an error term and i subscripts individual simulations. Detailed information on the operationalisation of this general model using simulations of Doha Round outcomes is provided in the following sections.

3. The literature sample

3.1 Data collection

Meta-analysts who wish to measure an average effect across studies (e.g. the effect of a new medicine tested on small samples of patients in different countries) strive to consider the entire literature on this effect. When the focus is placed instead on the comparison of typical methodological approaches, as in the case of applied trade models, a representative sample of the literature of interest is as valuable as entire literature, and much easier to collect.

We sample the literature on applied general and partial equilibrium simulations of DDR agricultural trade liberalisation. The aim is to sample publications in refereed journals but also in the so-called 'grey literature' where we suspect that an important share of the relevant simulations are published. We adopt the following sampling strategy:

From a recent review of trade models (van Tongeren, van Meijl and Surry 2001) we derive a vector of keywords that describe the type of applied models that we are interested in. We exclude stylized-numerical models as well as purely econometric models from the sample. Furthermore, we restrict the set of applications using a second vector of keywords taken from §13 of the Doha Ministerial Declaration (the section on agricultural markets). The combination of both vectors yields a matrix of search words that we apply to the most important literature databases (e.g. Econlit, Repec, etc.).[2] In addition, we search the internet in order to sample 'grey' literature that might not be listed in scientific literature databases.[3] Figure 1 presents information on the year of publication of the sampled studies and the average reported welfare change[4] over all sampled studies in each year from 1994 to 2005.

[Figure 1 here]

In our sample, the number of publications peaks twice in conjunction with the Doha and Cancun ministerial meetings (MM). Since the Cancun MM, the number of published simulations has fallen as expectations have been lowered and completely new proposals have not been forthcoming. Some authors have pointed to the 'shrinking gains from Doha' (Achterbosch,

[2] The keyword matrix and the literature databases employed are described in detail under http://memo-agecon.uni-goettingen.de/memo/searchstrategy.html. We are grateful to Inken Köhler, a trained librarian, for her assistance.
[3] Hess and Köhler (2005) provide general guidelines for literature searches in agricultural economics, and a list of related links can be found under http://www.uni-goettingen.de/en/sh/29498.html.
[4] We provide more information on the definition and measurement of this variable in the following section.

Hammouda, Osakwe and van Tongeren 2004 p.53; Ackerman 2005), addressing the fact that over time the reported findings from trade models have declined. Our literature sample supports this view, but only up to the Cancun MM (Figure 1). After Cancun, the average simulated gains from trade liberalization increase again. According to our analysis these movements in average simulated welfare changes are primarily due to the different proposals that were being considered in each phase of the DDR. In the years preceding the Doha MM, no concrete scenario was on the table and economists were mainly exploring the potential gains from often quite ambitious hypothetical scenarios.

3.2 Data processing and variable definitions

To estimate equation (1), the variables in it must be defined precisely. Our literature sample includes published articles, book chapters and proceedings of scientific conferences in addition to internet publications such as government reports and working papers. As a result, the quality of documentation with regard to variables in (1) varies widely. We find that studies that have been subject to a scientific review process are generally more transparently and thoroughly documented than others. In some instances, documentation of even very fundamental characteristics of a liberalisation experiment and/or the model used to produce simulation results is missing, and this forced us to omit some studies from the sample.[5]

The dependent variable (*I*) is defined as the simulated change in economic well-being (welfare) in a particular country/region due to a liberalisation step for a particular product/product aggregate. Most studies report, among other findings, some measure of welfare change, making this a natural choice to measure the impact of a liberalisation experiment. However, of 205 studies in our literature sample, a number of partial equilibrium (PE, e.g. FAPRI) and general

[5] We attempted to employ an internet questionnaire to collect missing information from the authors of studies in our sample. However, in a pre-test the response rate was very low. This is partly because some authors were not involved in the practical modelling underlying their studies and are therefore not able to answer our questions, while the actual ‘modellers’ can no longer be reached. Another problem with the questionnaire is the fact that different terminology is used by different ‘schools’ of PE and GE modellers, necessitating time-consuming clarification. Hence, we have stopped contacting authors, at least for the moment.

equilibrium (GE) applications report only changes in prices and/or trade volumes or related measures. Therefore, the literature sample is reduced to 122 publications. These publications report different welfare measures (equivalent variation, compensating variation, change in GDP, etc.; see Mas-Colell, Whinston and Green (1995) for a discussion of the relations between these measures). We transform all into million US$ and include dummy variables in (1) to account for differences in the measure used.

Quantifying the liberalisation experiments (*LE*) in a consistent and comparable manner is difficult. Much confusion about differences in simulation results may arise because important differences in what appear to be identical liberalisation experiments are not adequately considered. One difficulty is that the authors of simulation studies frequently describe their experiments in terms of rather broad *ad valorem* reductions of applied or bound protection but do not document the baseline levels of this protection that underlie their simulations. While two experiments that both simulate for instance a 50% cut in OECD agricultural tariffs may appear to be identical, they can differ considerably depending on the type of tariff used (bound versus applied, treatment of preferences and mixed tariffs, etc.) and the aggregation of sectors and regions in each model. The domestic price of good i in region r following a change in import tariffs can be given as:

$$p_{ri}^{\text{domestic}} = p_{ri}^{\text{border}} * (1 + t_{ri} + t_{ri}*s), \qquad (2)$$

where p_{ri} = the price of good i in region r, t_{ri} = the *ad valorem* tariff levied on imports of i, and s = the simulated proportional change in t_{ri}, which will in most instances be negative. Clearly, even moderate differences in t_{ri} paired with different levels of p_{ri}^{border}, due for example to different aggregations and treatments of tariffs, will lead to different impacts of a given s. Comparing simulation experiments is therefore not possible without controlling for the level of protection that has effectively been reduced in an experiment, and the economic size of the sector to which a tariff reduction s is applied.

To deal with this problem, we construct a reference database which includes information on tariffs, production volumes and trade flows. Key sources for this reference database are GTAP-5, FAO, and MacMap. With this reference database, we are able to approximately re-aggregate the regional and sectoral settings of any simulation experiment, and thus to calculate comparable initial *ad valorem* tariffs for each combination of country/region and products/product aggregate in a given study.[6] This *ad valorem* tariff is then multiplied with the proportional tariff change *s* and the average production value of the product in question (in million US$). The result is a standardised measure of the size of the liberalisation step underlying a particular simulation result (Figure 2).

This approach to operationalising the variable *LE* works for experiments that involve unilateral liberalisation for a given product/product aggregate and for a given country/region vis-à-vis the rest of the world. It does not, however, allow us to adequately measure the magnitude of the bilateral liberalization that occurs when, for example, two countries open their grain markets to one another exclusively. Hence, our method does not yet capture studies that analyze regional free trade agreements (FTAs) in a way that permits meaningful comparison. This is likely not a serious shortcoming because although the “New Bilateralism” has been a recurrent issue in course of the DDR, few studies in our sample focus primarily on FTAs. This may be due to our search strategy that concentrated on global agricultural trade liberalization scenarios.

A further issue concerns the liberalization that is implicit when a database is transformed to some projection year. For example, when authors project the GTAP-5 database from its base year 1997 to the year 2005 as a starting year for a policy simulation, we include the policy changes that they ‘build’ into this projection as part of the full policy experiment in order to approximate the true magnitude of the effective reduction in protection relative to a base year.

[Figure 2 here]

[6] We are grateful to Tinoush Jamali for programming assistance.

A variety of variables are defined and measured to capture model characteristics (*MC*). These include: whether the model is PE or GE; the closure used in GE cases; whether the Armington assumption is employed and, if so, the magnitude of the Armington elasticities; whether constant or increasing economies of scale are assumed; how aggregated the model is in terms of countries/regions and products; etc. As with *LE*, the quantification of model characteristics (*MC*) depends on the quality of the documentation in the studies in our sample. The importance of data quality cannot be underestimated in this regard. Harrison et al. (1997), for example, claim that the effects of increasing returns to scale in the model used in studies by Francois (compare e.g. Francois, van Meijl and van Tongeren 2003) are largely due to increased elasticities of substitution between primary and intermediate inputs. We were able to collect information on these elasticities in the models employed by both Harrison et al. and Francois, but for many other models in our literature sample these elasticities could not be determined. Therefore, we cannot (yet) test this issue econometrically and are reduced to capturing differences in economies of scale between studies with a dummy variable.[7]

Adequate coverage of dynamic simulations is also difficult. Our method only captures average welfare changes relative to a base year, omitting the information that dynamic models generate about the paths taken by these changes over the time horizon of a simulation run.

Databases (*DB*) are incorporated in the form of dummy variables for the GTAP-4 and GTAP-5 databases, as well as dummies for other databases from before and after 1997. The dummy variable for GTAP-3 is excluded, so all database effects are measured relative to GTAP-3. There is only one publication employing GTAP-6 in our literature sample (sample ends in 9/2005), so it is not possible to measure this effect separately.

To capture study characteristics (*SC*), many meta-analyses use descriptive bibliographical information about studies (year of publication, number of authors) as well as information about the contexts in which the studies have been conducted (institution of origin, subject to peer

[7] In an open version of our survey (http://memo-agecon.uni-goettingen.de/phpsurveyor/index.php?sid=8), we invite those interested to point us towards specific publications and post details about models and simulation designs.

review, etc.). *SC* variables that we have captured and tested include: i) the type of institution at which a study has been conducted (i.e. university vs. international institution); ii) year of publication; iii) whether or not the study underwent scientific review; and iv) the number of authors per study. However, we do not include these variables in the final estimation of equation (1) because they fail to have a significant impact. We suspect that this is because these variables, if they do influence the results of a study, will do so indirectly via model characteristics. For example, if an institution that sponsors a study is interested in large simulated welfare changes to underscore its case for liberalisation, this will probably be reflected in a choice of liberalisation experiments (*LE*) and model characteristics (*MC*) that favours the desired results. Of course, we have no way of knowing the extent to which a researcher's *LE* and *MC* choices are motivated by the conviction that these choices are objectively superior, or by a desire to generate specific results, and we do not mean to suggest that researchers knowingly manipulate results.

Figure 3 provides an overview of the dataset generation system outlined above. This system was automated using MySQL, a freely available, open-source database software, and can be operated from an MS-Access interface.

[Figure 3 here]

3.3 Special cases

Publications using the Michigan Model of World Production and Trade (by Brown, Deardoff, Stern, et al., BDS) for some liberalization scenarios report simulated welfare gains for specific countries (as well as for the sum of certain experiments) that are 50% to 100% larger than the average gains of all other studies. It is well known that this model generates the highest absolute welfare gains that are produced by contemporary trade models. Initial estimates of equation (1) attributes these large gains to the fact that the Michigan model uses the Johansen macro closure, keeps trade balances for all countries fixed, explicitly models flows of foreign direct investment

(FDI) as a result of liberalization, uses own estimates of firm level markups (for imperfect competition), and partly includes own estimates of non-tariff barriers (NTBs) that are removed along with tariffs. In addition, this model has a Rest of the World region against which all other prices rise and for which BDS do not report results in the studies included in our literature sample. Only one other model in the literature sample ('China WTO' by Wang) uses the Johansen closure as well, accompanied by FDI liberalization, but without specific estimates of NTBs. This model also predicts comparatively large gains for China and some other countries in a WTO accession simulation under assumptions of constant or increasing returns to scale.

Our reference database for calculating standardised liberalisation experiments (*LE*) cannot deal with publication-specific estimates of NTBs. Nor have we been able to quantify flows of FDI in a way that would make these effects comparable across studies. At any rate, most models that focus on agricultural trade liberalization and changes in agricultural policies do not employ these features. The leverage (Cook's distance) of these features (proxied with dummy variables) on the estimation of equation (1) is extraordinary large. We therefore exclude the BDS and China WTO studies from our final dataset. Instead, we concentrate on those simulation experiments which do not explicitly exclude agriculture, do not emphasise reductions of NTBs based on own estimates, and do not model the liberalisation of FDI flows in a way that is not transparent. Note that the estimation of equation (1) with the complete dataset (including BDS, and studies that model FDI and NTBs) leads to the same general findings that are reported below; however, the overall fit of the model is lower (R^2 drops from 77% to 48%).

The exclusion of the BDS and similar studies, as well as other exclusions due to inadequate documentation, etc. reduce the number of studies used to produce the estimates of equation (1) reported below to 53. These 53 studies contain a total of roughly 1600 individual observations, each representing the impact of a particular liberalisation of specific products/product aggregates summed for a specific country/region. Our reference system captures welfare changes at the product level and would thus enable far more disaggregated comparisons. However, almost all publications report results at the country level and therefore we have chosen this aggregation level for our dependent variable.

4. Estimation and results

Equation (1) is estimated using OLS. We recognize econometric difficulties that arise from the complex, non-constant variance structure of the dataset employed, and anticipate that refinement of the estimation technique will generate additional insights. The selection of independent variables and interaction terms in the linear model is accomplished according to a stepwise regression algorithm so as not exclude effects that are unanticipated *a priori*. This entirely empirical approach is one of the strengths of meta-analyis, because any hypothesis derived from theory can be tested, but the data are permitted to reveal other, even stronger effects. Results of the estimation of equation (1) are presented in Table 1.

[Table 1 here]

The independent variables that are included in the estimation results in Table 1 are not the only ones that have significant influences on the simulation results. However, many of the variables in the *MC* and *DB* vectors are correlated with one another. It is therefore important to choose variables that are as independent of each other as possible (Florax 2002), and at the same time are not by definition restricted to only a subset of the models in our dataset. An example in this regard are different assumptions about factor mobility in CGE models: since many PE models do not depict factors such as labor or capital, measurement of these effects would only be possible for a regression on a subset of CGE studies. We therefore do not consider these variables. We discuss the main results in the following.

<u>Policy shocks</u>: For each percent of effective reduction in tariff protection (including export subsidies and amber box support, which we approximate through the tariff level), welfare gains

amount to US$ 21,000 per million US$ of production value. Gains from production-related blue and green box reductions are also statistically significant but much smaller, partly because some partial experiments model the EU's recent decoupling reforms as reductions of direct payments and a simultaneous increase in decoupled payments.

Shocks to technical change parameters: In a few studies, simulations provide insight into the effect of shocks to technical change, e.g. when genetically modified plants are adopted by farmers. We have expressed these effects relative to the production value of a sector. The estimated coefficient states that a one percent shock to this variable yields on average 1.6 million of welfare gains. This huge average gain from technology improvements and benefits from agricultural research and extension is not implausible relative to observed and estimated gains from trade liberalization alone. In their meta-analysis, Alston, Chang-Kang, Marra, Pardey and Wyatt (2000) find an average 100% rate of return to agricultural research and development.

Databases: The coefficients for GTAP-5 and other databases prior to 1997 are significant and negative, suggesting that the use of these databases leads *ceteris paribus* to lower simulated welfare gains. The documentation for GTAP-5 mentions it was constructed using high 1995 world market prices for agricultural products. Our results contradict findings in the literature that attribute large simulated welfare gains to GTAP-5 and significantly lower gains to GTAP-4 and GTAP-6. For randomly selected agricultural products and countries we observe differences of 50% to 100% in production values between databases (GTAP-5, GTAP-6, the ATPSM database and our own reference based on FAO data averaged from 1995 to 2000). We conclude that databases, although at the core of each modelling exercise, are far from being homogeneous, and that different base years can have a strong influence on simulation results. More research and perhaps a global data gathering effort are required to understand and deal with this effect.

Dynamic model: This dummy must be interpreted in connection with the 'fixed capital stock' dummy. The latter captures the difference between the standard comparative static GE model with fixed stocks for capital and labor that do not accumulate during a simulation period. This is in contrast to PE models, which usually do not consider any capital restrictions (except those implicit in the supply elasticities), and GE models with any form of capital accumulating closure. Dynamic models are considered within our framework as GE models of the latter kind since one of their distinct features is that the capital is permitted to grow during a simulation run. At the same time there are PE models with a recursive dynamic structure that do not include capital at all. Controlling for the capital accumulation effect, dynamic models predict larger welfare gains to liberalisation than static models, but this effect is not significant. This is in line with the qualitative findings reported by Brown, Kiyota and Stern (2004) (who also cite Harrison, Rutherford and Tarr (2003)).

Disaggregation of the agricultural sector: Increased disaggregation of the agricultural sector is associated with smaller welfare losses, which can be explained by the fact that some aggregations in CGE models are vertical rather than horizontal combinations of agricultural sectors across all levels of processing. Thus, comparatively large artificial sectors such as 'livestock and livestock products' are created in some models, and this seems to be associated with much larger gains than more 'natural' aggregations for similar levels of processing.[8]

Armington elasticities: Large Armington elasticities (values that are twice those in the standard GTAP model or higher) lead to significantly higher simulated welfare gains. Note that some partial models also employ the Armington assumption, but not always through a CES specification (e.g. the Armington feature of ATPSM just 'simulates' Armington behaviour).

Partial vs. general equilibrium: Controlling for the size of the liberalisation step and all other explanatory factors included in Table 1, welfare changes simulated using PE models are on average 2.3 billion US$ higher than the changes simulated using GE models (excluding BDS). Among the GE models, the impact of modelling increasing returns to scale in some sectors (typically not primary agriculture) appears to be insignificant. However, the interaction between

[8] In fact, we find large differences between publications with regard to what they consider as 'agriculture'. A useful quantification of these different treatments is possible using our reference database, but this has not yet been incorporated into our meta-analysis.

‘increasing returns to scale’ and ‘trade volume’ reveals that large traders tend to loose under such scenarios. This is in line with findings in the literature according to which increasing returns to scale are most important in developing countries and particularly increase their competitive advantage in trade if they can be captured.

5. Discussion and conclusions

According to Florax, de Groot and de Mooij (2002 p. 5, 12), “(d)oing a credible meta-analysis is not something that is done in a rainy Sunday afternoon… In order to ensure usefulness for applied general equilibrium modelling, a close collaboration between theorists and empirical researchers is desirable.” These authors also state that “although meta-analysis … necessitates the time-consuming construction of databases, there is an indisputable pay-off in terms of obtaining improved ‘consensus estimates’ and detailed insights in the available empirical material.”

The results presented above suggest that a simple linear regression using variables that describe the liberalisation experiment, the characteristics of the model, and the database employed can identify important causal relationships that are plausible and explain a major share of the variance in the dependent variable ‘welfare change’ in a sample of Doha Round trade liberalisation studies. In this sense, our analysis demonstrates that meta-analysis can be combined with applied trade modelling in a way that casts light on the impact of key model characteristics such whether a model is PE or GE.

This does not mean that our simple meta-regression can sufficiently approximate the findings of simulation models for any specific liberalisation experiment, in the sense of a ‘meta-model response surface’. Examination of the results of our meta-analysis reveals a significant number of outliers. Many of these appear to be associated with terms of trade effects, which we have not succeeded in capturing adequately so far. The quantification of liberalisation experiments and a number of important model characteristics in our dataset could certainly be refined further, and the underlying literature sample could be expanded, for example to include trade simulation models that use the gravity approach. Finally, the use of more sophisticated econometric techniques will likely lead to important refinements in our results.

Our results may not yield many new insights for modelling experts. However, for policymakers and all others on the demand side of applied trade modelling we conclude that:

- Trade simulation models are comparable and can generate quantitative insights into real world policies if the combination of assumptions behind a simulation experiment is carefully controlled for and understood.
- A comparatively small set of fundamental model characteristics can explain the majority of the variation in simulation results across studies of DDR agricultural liberalisation. For example, we find that partial equilibrium models produce significantly larger estimates of welfare gains than general equilibrium models, *ceteris paribus*.
- The causality behind applied trade models is complex. A variety of model and database characteristics interact to determine the results of a liberalisations experiment, and general conclusions about the influence of individual components cannot readily be drawn without detailed empirical investigation.
- Our estimates may help researchers and policy makers to focus on the most suitable modelling assumptions for specific policy problems rather than to devote effort to the development of a single model that suits all possible questions.

6. References[9]

Achterbosch, T. J., Hammouda, H. B., Osakwe, P. N., van Tongeren, F. W., 2004. Trade liberalisation under the Doha Development Agenda. The Hague, Agricultural Economics Research Institute (LEI), Report 6.04.09.

Ackerman, F., 2005. The Shrinking Gains from Trade: A Critical Assessment of Doha Round Projections, Global Development and Environment Institute 05,01.

Alston, J. M., Carter, C. A., Green, R., Pick, D., 1990. Whither Armington trade models?, American Journal of Agricultural Economics 72, 455-467.

Alston, J. M., Chang-Kang, C., Marra, M. C., Pardey, P. G., Wyatt, T., 2000. A Meta-Analysis of Rates of Return to Agricultural R&D - Ex Pede Herculem? Washington D.C., International Food and Policy Research Institute.

Alston, J. M., Norton, G. W., Pardey, P. G. 1995. Science under scarcity: principles and practice for agricultural research evaluation and priority setting, Ithaca [u.a.], Cornell Univ. Press.

Anderson, J. E., Wincoop, E. v., 2001. Borders, Trade and Welfare, Nber Working Paper Series 8515.

Brown, D. K., Kiyota, K., Stern, R. M., 2004. Computational analysis of the U.S. bilateral free trade agreement with Central America, Australia, and Marocco, Gerald R. Ford School of Public Policy, The University of Michigan, Ann Arbor, Michigan 48109-1220, Discussion Paper No. 527.

Chalmers, I., Altman, D. G., (Eds). 1995. Systematic Reviews, BMJ Publishing Group.

[9] A list of the publications included in our literature sample is available under http://memo-agecon.uni-goettingen.de/memo/literaturesample.html.

Charlton, A. H., Stiglitz, J. E., 2005. A Development-friendly Priorisation of Doha Round Proposals, The World Economy 28,3, 293-312.

Devarajan, S., Robinson, S., 2002. The Influence of Computable General Equilibrium Models on Policy, International Food Policy Research Institute.

Eysenck, H. J. 1995. Problems with Meta-Analysis. In: Chalmers, I. Altman, D. G. (Eds), Systematic Reviews, BMJ Publishing Group.

Florax, R. J. C. M., 2002. Accounting for Dependence among Study Results in Meta-Analysis: Methodology and applications to the Valuation and Use of Natural Resources, Serie Research Memoranda: Research Memorandum 2002-5 Vrije Universiteit Amsterdam.

Florax, R. J. G. M., de Groot, H. L. F., de Mooij, R. A., 2002. Meta-analysis, Tinbergen Institute Discussion Paper 2002-041/3.

Francois, J., van Meijl, H., van Tongeren, F., 2003. Economic Implications of Trade Liberalization Under the Doha Round, GTAP Technical Paper, GTAP website download 10.6.2003.

Gohin, A., Moschini, G., 2005. Evaluating the Market and Welfare Impacts of Agricultural Policies in Developed Countries: Comparison of Partial and General Equilibrium Measures, Review of Agricultural Economics 28,2, 195-211.

Harrison, G. W., Rutherford, T. F., Tarr, D. G., 1997. Quantifying the Uruguay Round, The Economic Journal 107,444, 1405-1430.

Harrison, G. W., Rutherford, T. F., Tarr, D. G., 2003. Rules of Thumb for Evaluating Preferential Trading Arrangements: Evidence from Computable General Equilibrium Assessments, Working Paper WPS# 3149 (October).

Hertel, T, Hummels, D., Ivanic, M., Keeney, R., 2003. How Confident Can We Be in CGE-Based Assessments of Free Trade Agreements?, GTAP Working Paper No. 26.

Laird, S., Cernat, L., Turrini, A., 2003. Back to Basics: Market Issues in the Doha Agenda. New York and Geneva, UNCTAD.

Mas-Colell, A., Whinston, M. D., Green, J. R. 1995. Microeconomic Theory, New York, Oxford University Press.

McKitrick, R. R., 1998. The Econometric Critique of Computable General Equilibrium Modeling: The Role of Parameter Estimation, Economic Modeling 15,543-573.

Piermartini, R., Teh, R., 2005. Demystifying Modelling Methods for Trade Policy, World Trade Organization Discussion Paper No. 10.

Robinson, S. 1989. Multisectoral Models. In: Chenery, H. Srinivasan, T. N. (Eds), Handbook of Development Economics Vol. II, Elsevier.

Scollay, R., Gilbert, J., 2000. Measuring the Gains from APEC Trade Liberalization: An Overview of CGE Assessments, The World Economy 23,2, 175-197.

Stanley, T. D., 2001. Wheat from Chaff: Meta-Analysis As Quantitative Literature Review, Journal of Economic Perspectives 15,3, 131-150.

Stanley, T. D., 2005. Beyond Publication Bias, Journal of Economic Surveys 19,3, 309-345.

Stanley, T. D., Jarrel, S. B., 1989. Meta-Regression Analysis: A Quantitative Method of Literature Surveys, Journal of Economic Surveys 3, p. 161-170.

van Tongeren, F., van Meijl, H., Surry, Y., 2001. Global Models Applied to Agricultural and Trade Policies: A Review and Assessment, Agricultural Economics 26, 149-172.

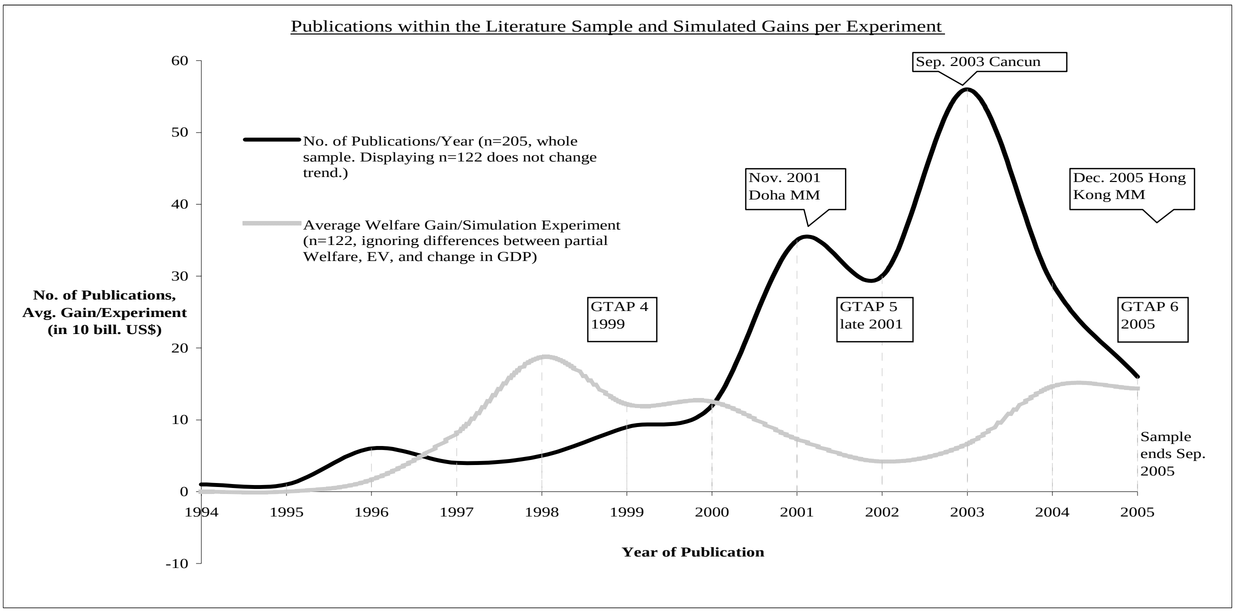
Wachter, K. W., 1988. Disturbed by Meta-Analysis, Science 241, p. 1407-1408.

Table 1: OLS regression results for the restricted literature sample (weighted by number of observations from each publication)

Variables	Mean	Standard deviation	Minimum	Maximum	Estmated coefficient	Standard error	t-value	Prob(>|t|)
Intercept					2034.6	917.3	2.22	0.026
Database GTAP-4 (dummy = 1 if yes)	0.24		0	1	-1068.2	933.3	1.14	0.252
Database GTAP-5 (dummy = 1 if yes)	0.33		0	1	-1549.8	912.6	1.70	0.089
Other database after 1997 (includes GTAP-6) (dummy = 1 if yes)	0.30		0	1	327.1	1043.1	0.31	0.753
Other database before 1997 (dummy = 1 if yes)	0.07		0	1	-3299.4	960.5	3.44	0.000
Number of regions (count)	55.5	63.99	1	161	-7.4	5. 9	1.26	0.206
Number of agricultural sectors (count)	16.47	13.17	1	40	-72.1	24.2	2.98	0.002
Trade volume of country (million US$)	175912	302211	2.247	2928459	0.015	0.0004	32.17	0.000
Shocks to technical change or related variables in per cent (million US$ / 1% shock)	28.4	804	0	22807	1.6	0.25	6.24	0.000
Changes in tariffs, export subsidies and amber box measures (million US$ / 1% reduction in protection)	-39260	190811	-4503000	35270	-0.021	0.0008	25.42	0.000
Changes in blue/green box policies (million US$ / 1% policy shock)	-128400	1625902	-30930000	31660	-0.0002	0.00008	2.24	0.025
Armington with high elasticities (dummy = yes for twice GTAP standard or higher)	0.18		0	1	2282.4	484.3	4.71	0.000
Increasing returns to scale (dummy = yes if IRTS in some sectors)	0.12		0	1	403.4	901.0	0.45	0.654
Trade volume * increasing returns to scale (interaction term)					-0.01	0.002	4.98	0.000
Dynamic model (dummy = 1 if yes)	0.21		0	1	506.7	457.4	1.11	0.268
Fixed capital stock (dummy = 1 if yes)	0.44		0	1	-1378.3	460.4	2.99	0.002
Partial equilibrium model (dummy = 1 if yes)	0.33		0	1	2263.3	6781	3.34	0.000

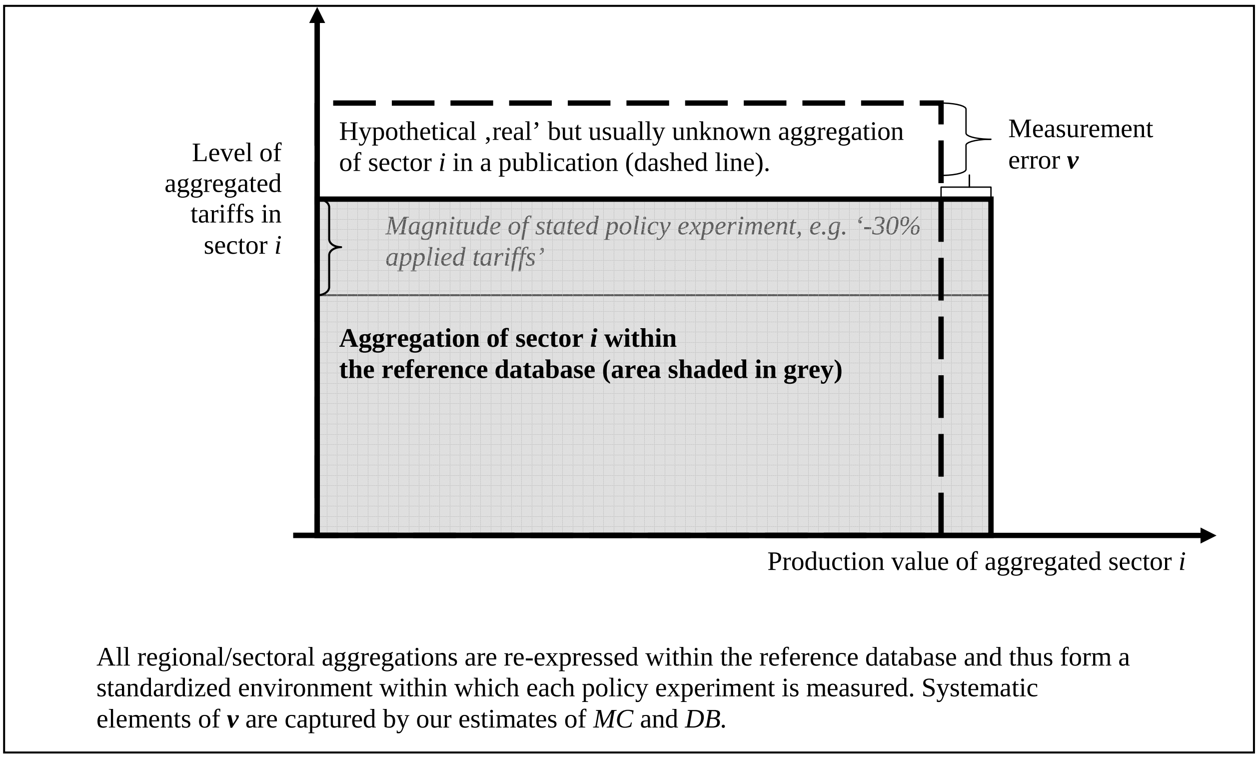
Residual standard error: 1150 with 1587 degrees of freedom.
Multiple $R^2 = 0.773$, Adjusted $R^2 = 0.771$.
F-statistic: 338 with 16 and 1587 degrees of freedom, p-value = 0.000.
Source: Own

Figure 1: Number of publications in our literature sample, and average reported welfare gain (1994-2005)



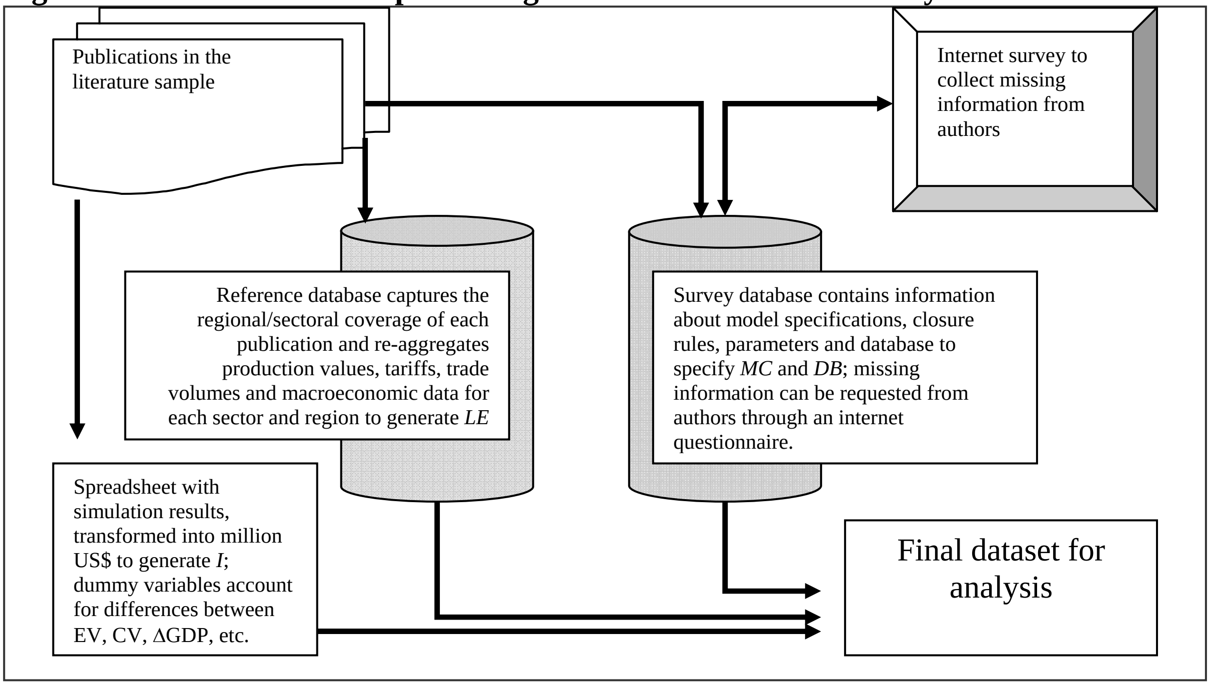
Source: Own calculations based on literature search.
Notes: 83 of 205 studies do not report welfare changes. MM = ministerial meeting.

Figure 2: Calculating a standardized measure of the liberalisation experiment underlying a simulation



Source: Own presentation.

Figure 3: The collection and processing of the dataset for meta-analysis



Source: Own presentation.