CATPRN
Canadian Agricultural Trade Policy Research Network

AGRICULTURAL TRADE POLICY MODELLING: INSIGHTS FROM A META-ANALYSIS OF DOHA DEVELOPMENT AGENDA OUTCOMES

CATPRN Commissioned Paper CP 2008-02
May 2008

Sebastian Hess
Post-Doctoral Candidate

Stephan von Cramon-Taubadel
Professor

Department for Agricultural Economics and Rural Development
University of Göttingen

http://www.catrade.org

Funding for this project was provided by the German Research Foundation the Fullbright Commission and the Canadian Agricultural Trade Policy Research Network (CATPRN). The views expressed in this paper are those of the authors and should not be attributed to the funding organizations. The authors appreciate the helpful comments and suggestions provided by Karl Meilke, Martina Brockmeier, Bernhard Brümmer and Yves Surry. All remaining errors and omissions are our own.

Abstract

In a meta-analysis of trade policy models, Hess and von Cramon-Taubadel (2008) use over 5800 simulated welfare effects from 110 studies of potential Doha Development Agenda outcomes to identify characteristics of models, data and policy experiments that influence simulation results. This meta-analysis, which is recapitulated here, produces plausible results and explains a significant proportion of the variation in simulated welfare effects. However, due to insufficient documentation and the complexity of the general and partial equilibrium models in the literature sample, many explanatory variables employed in this analysis are binary. This precludes more detailed analysis of their impacts across models. Therefore, a partial equilibrium model and a single country CGE for Canada are employed to generate synthetic meta-data. Simulation scenarios are based on random combinations of base data, elasticities and tariff changes selected from plausible ranges obtained from the literature sample. The synthetic meta-data has the advantage that the values of explanatory variables are measured exactly. This makes it possible to explore more complex issues of functional form and interaction between variables in the meta-analysis. The results indicate for both models that first- and second-order polynomials provide sufficient approximations of the model response. Especially in the CGE model, interaction terms between elasticities and policy variables are important. We conclude that meta-analysis can provide insights into the behaviour of trade policy models beyond what is possible with conventional sensitivity analysis and qualitative reviews.

1.0 INTRODUCTION

Economists employ applied trade models to generate empirical estimates of the gains and losses that would result from trade liberalisation. Model-based predictions of significant global and national gains from trade liberalisation are cited to support the case for a successful conclusion of the Doha Development Agenda (DDA) and other trade negotiations (e.g. Anderson et al., 2006). However, different models often produce trade simulation results that "differ quite widely even across similar experiments" (Charlton and Stiglitz, 2005). Experienced modellers can offer explanations for these differences, but these explanations are largely partial and often quite technical. For many users of trade model simulations, variation in model simulation results is confusing, and most proposed explanations are inaccessible. This complicates an already controversial debate on trade liberalisation. It confirms the doubts of critics who question the ability of economists to accurately estimate the benefits of liberalisation, or who question the existence of these benefits in the first place. At the limit, the credibility of a large and visible branch of the economic profession is at stake.

Against this background, Hess and von Cramon-Taubadel (2008) investigate whether meta-analysis can contribute to explaining variation in quantitative trade policy model simulations. Meta-analysis has been used in a number of different ways in the trade modelling literature, typically to provide reliable estimates of the elasticities and other parameters that are a vital component of quantitative trade models (Boys and Florax, 2007). In this paper, however, meta-analysis is used to identify model characteristics (e.g. partial vs. general equilibrium specification, constant or increasing returns to scale) and other factors (e.g. the database employed, the size of the simulated liberalisation step) that influence simulation results in a systematic manner, and to derive quantitative estimates of these influences.

To this end, and following a discussion of our conceptual approach in section 2, we present the results of two meta-analyses. First, in section 3 we briefly summarise the main results of Hess and von Cramon-Taubadel (2008), who present a meta-analysis of published DDA outcomes (roughly 5800 simulated welfare effects from 110 studies). A discussion of the strengths and weaknesses of this first literature-based meta-analysis motivates the second meta-analysis in section 4, which is based on a synthetic dataset that we generate using two 'typical' models, one partial equilibrium (PE) and the other general equilibrium (GE). As discussed below, this second meta-analysis can be interpreted as an extensive sensitivity analysis, but it is in fact a form of response surface analysis. Section 5 concludes.

2.0 CONCEPTUAL APPROACH[1]

2.1 Why do the results of quantitative trade policy simulations vary?

Based on a review of the literature on quantitative trade models, Hess and von Cramon-Taubadel (2008) identify five categories of factors that can lead to variation in the results of trade model simulations. First, different studies measure different outcomes (e.g. changes in welfare, changes in GDP, etc.). Second, liberalisation scenarios differ across studies. Third, different studies are based on different model specifications (e.g. partial vs. general equilibrium, constant vs. increasing returns to scale, etc.). Fourth, studies are based on different datasets (e.g. GTAP-4 vs. GTAP-6). Finally, what might be termed a study's 'research context' (e.g. the authors' affiliation, whether a study was subject to peer review) might also influence its results.

[1] This section and section 3 draw on Hess and von Cramon-Taubadel (2008).

The results of any trade model simulation will be a complex function of the many factors in these five categories, and of interactions between them. Economists attempt to measure the impact of these factors using sensitivity analysis and qualitative reviews. Sensitivity analysis is typically used to cast light on factors in the second and third categories listed above. However, conventional sensitivity analyses rarely vary more than a few parameters simultaneously. Under these circumstances, conventional sensitivity analysis is unable to capture the multivariate complexity underlying a trade model simulation. In addition, and as a result, conventional sensitivity analysis is not well-suited to comparing simulation results across models. Qualitative reviews of published studies (Charlton and Stiglitz, 2005; Piermartini and Teh, 2005) have been used to compare results across models, typically grouping them according to selected model characteristics (e.g. 'dynamic vs. static'), or types of liberalisation experiment. However, such comparisons are also unable to control for simultaneous variation in the other many factors listed above.

To overcome these limitations of sensitivity analysis and qualitative review, and thus to generate deeper insights into the results of quantitative trade policy simulations, a new approach is needed. This approach should permit the comparison of trade model simulations while controlling for simultaneous variation in the five types of factors listed above. In this paper we explore whether meta-analysis is suited to this task.

2.2 Meta-analysis

Meta-analysis is an inductive empirical approach used to identify similarities and explain differences between scientific findings across studies (Stanley, 2001). When meta-analysis was developed in the fields of medicine and psychology, the primary goal was that of combining evidence. Since experiments with human beings are often based on small samples, combining experiments to create a meta-sample can lead to more precise estimation of treatment/effect relations. This type of meta-analysis has also been employed in the trade policy modelling literature to derive reliable estimates of key model parameters such as elasticities (Boys and Florax, 2007). Another use of meta-analysis, and the one illustrated here, is in the evaluation of methods. This type of meta-analysis is used to quantify the share of variance within a given set of estimates that is due to variation in methodologies, assumptions and other factors. This approach has evolved especially in disciplines such as economics in which reproducible measurements are often hard to obtain and quantitative results are known to depend heavily on the methods that are applied (Stanley, 2001).

We wish to explore whether meta-analysis has the potential to add to the insights provided by sensitivity analysis and qualitative reviews by permitting quantitative comparison of trade model simulations that controls for simultaneous variations in measured effects, model characteristics, liberalisation steps, databases and research contexts. The basic model underlying the meta-analyses presented here is therefore:

$$I = f(MC, LE, DB, RC, u), \tag{1}$$

where I is the simulated impact of a trade liberalisation experiment, MC is a vector of model characteristics (e.g. PE or GE, depiction of returns to scale), LE describes the liberalisation experiment (e.g. the magnitude of the simulated tariff reduction), DB is the database underlying the simulation (e.g. GTAP-4 or GTAP-5), RC is a vector of research context variables (e.g. affiliation of the authors, whether the study has been published), and u is an error term. In the following sections we estimate this model first using a dataset extracted from published simulations of DDA liberalisation outcomes (section 3), and second using a synthetic dataset generated using typical PE and GE models (section 4).

3.0 META-ANALYSIS OF PUBLISHED DDA LIBERALISATION SIMULATIONS

3.1 Sample collection

Hess and von Cramon-Taubadel (2008) outline a sample collection strategy that generates a representative initial raw dataset comprising over 1200 DDA liberalisation simulations. Setting aside (i) studies that do not present original simulations but rather draw on the results of earlier simulation studies, (ii) redundancies (sometimes the same simulation results appear in a working paper, a conference paper and a journal publication), (iii) studies that provide no information on the model used to produce a simulation (for example, some studies only make reference to 'World Bank estimates' or 'GTAP'), (iv) studies that do not report results in numerical form (some studies only present graphs), and (v) studies that do not report welfare changes (but rather, for example, changes in trade flows or production volumes) leaves 230 studies from the years 1994 to 2006.[2]

Of these 230 studies, roughly 60 studies are omitted because on closer examination they do not focus on DDA trade liberalisation but rather on related but distinct topics such as regional free trade agreements or the poverty-alleviating effects of trade liberalization. A further 60 studies are not included in the final sample due to insufficient documentation. It turns out that in many instances documentation of even very fundamental characteristics of a liberalisation experiment and/or the model characteristics used to produce simulation results is missing. Of course, it is sometimes difficult for modellers to provide detailed documentation even if they wish to; limits on the length of journal articles, book chapters and conference papers frequently preclude full documentation of the data, parameters and assumptions that underlie a simulation exercise. Nevertheless, considering both those studies that are dropped from the original 1200 because they provide no information about the underlying model, and those that are omitted later on because the information that they do provide is found to be lacking in some key respect, a clear and disturbing conclusion is that incomplete documentation is common in the trade modelling literature.

Heterogeneous documentation creates a trade-off between (i) the number of studies included in the final meta-sample, and (ii) the number of independent variables that can be quantified and included in estimation of equation (1). As the number of variables to be included in the estimation increases, the set of studies that provide sufficient information on these variables shrinks. The final sample of 110 studies is the result of this trade-off.

3.2 Specification and variable definition

Each study in the sample presents the results of one or more liberalisation experiments (often called 'scenarios'). Each liberalisation experiment simulates welfare changes in one or more countries or aggregated geographic regions. In our meta-analysis, each individual country or region for which the impact of a liberalisation experiment is reported is considered a single observation. This definition of an observation has two important implications. First, since the average study simulates more than one liberalisation experiment, and since the average model depicts more than one country/region (and some PE models depict more than 100 countries), a literature sample that includes 110 studies generates many more than 110 observations for the estimation of equation (1). In fact, each of the 110 studies in the final sample produces just over 53 observations, for a total of 5835 observations in the meta-analysis.[3] Second, while some studies produce many observations, others produce as few as one (if they report on only

[2] A list of these studies is available from the authors on request.
[3] A list of the 110 studies in the final sample, and information on the number of observations produced by each study, is available from the authors on request.

one scenario for one country/region). As a result, studies will have different weights in the meta-dataset, and thus different influences on the econometric estimation of equation (1). For this reason, a weighted version of equation (1) in which each observation is divided by the number of observations produced by the underlying study is estimated. This weighting scheme gives each of the 110 studies in the final literature sample the same weight (Weichselbaumer and Winter-Ebmer, 2005; Knell and Stix, 2005; Stanley, 1998).[4]

In the estimation of equation (1), the dependent variable (I) is defined as the simulated economic welfare change in a particular country/region due to a liberalisation experiment in million 2001 US$. Since the countries/regions behind the individual observations in the dataset vary considerably in economic size, net trade position and geographic location, the welfare gains that they experience as a result of liberalisation will vary as well. To account for these differences we include a set of dummy variables that capture country/region fixed effects on the RHS of equation (1). Since some countries and regions appear in different aggregations (e.g. EU-12, -15 and -25), and a wide variety of groups is considered in different studies (G-20, Cairns, etc.), altogether 339 dummies are required to capture country/region fixed effects.

The specification of independent variables in the categories MC, LE, and DB is detailed in Hess and von Cramon-Taubadel (2008). In the category MC (model characteristics), 20 dummy variables are used to captures the effects of key model characteristics and interactions between them in general equilibrium (GE) and partial equilibrium (PE) applications. These characteristics include whether a model assumes increasing or decreasing returns to scale, the size of the Armington elasticities, whether the neoclassical or the Johansen closure is used (in GE applications) and whether the capital stock is fixed or accumulates. Other dummy variables reflect whether trade balances are assumed to be fixed in a simulation, and whether authors report having made any modifications to the elasticities that they have adopted from the literature or that are a part of the modelling platform employed. For dynamic models, the length of the simulation run in years is included in linear and quadratic form. Three variables measure the disaggregation of the model in terms of countries/regions, sectors and agricultural products, respectively.

The category LE (liberalization experiment) is difficult to quantify consistently. Much confusion about differences in simulation results arises because important differences between what appear to be identical liberalisation experiments are overlooked. Using a reference database that includes information on applied and bound tariffs, production volumes and trade flows, Hess and von Cramon-Taubadel (2008) construct a variable labelled 'changes in tariff protection' that provides a standardised monetary measure of the size of the liberalisation step underlying a particular simulation.[5] This variable is also included in quadratic form in the estimation of equation (1) to account for non-linearity in the relation between welfare effects and price wedges. Changes in export subsidies, export taxes, amber box measures, blue and green box policies and non tariff barriers (NTBs) are produced

[4] Hess and von Cramon-Taubadel (2008) present results for different sub-samples, with and without the weighting scheme. The main results are robust across specifications.

[5] This variable takes the value of the product that is subject to liberalisation and multiplies it first by the initial *ad valorem* applied tariff (to produce an estimate of the value of protection that is being changed) and second by the simulated percentage reduction in tariffs (to produce an estimate of the value of the liberalisation step). For example, consider a study that simulates a 20 percent reduction in EU oilseed tariffs. With the reference database we determine the aggregated value of EU oilseed production and the *ad valorem* tariff applied to EU oilseed imports in the study's base period. (Production value)*(applied *ad valorem* tariff)*(20 percent) provides a measure (in US$) of the change in EU oilseed protection in the liberalisation experiment.

analogously[6]. A final LE variable captures whether a simulation assumes an exogenous shock to technology or any related parameter that influences productivity in a model.

DB (database) variables are incorporated by means of dummy variables. GTAP-3 is the reference database, and one dummy variable each is included for GTAP-4, GTAP-5 and GTAP-6. Two dummies are included for 'other' databases (such as national account data for single country CGEs, or agricultural production and trade statistics for most PE models); one each for databases based on bound and applied tariffs.

Not many RC (research context) variables could be quantified, and those that could (year of publication, number of authors, subject to peer review) have no significant impact in estimates of equation (1). We conclude that RC influences will be largely captured by MC variables, i.e. authors who 'intend' to produce larger (or smaller) estimates of welfare gains from liberalization will do so by selecting – in the process of the "model pre-selection" discussed in Hertel (1999) – model characteristics that are expected to generate such estimates.

3.3 Results

Selected results of the estimation of equation (1) are presented in Table 1. We estimate a simple additive linear OLS specification, although some quadratic terms and interactions (the 20 model characteristic dummies) are also considered. The coefficient of determination (R^2) is 55.7 percent and estimates produced using different specifications and sub-samples (presented in detail in Hess and von Cramon-Taubadel, 2008) indicate that the main results of the literature based meta-analysis are robust.

Results for the 'change in tariffs' variable indicate that larger tariff reductions lead to larger welfare gains. These gains equal 45,400 US$ per 1 million US$ reduction in tariff protection. The squared tariff change term is positive and not significant, so we are unable to confirm the expected quadratic relation between tariff changes and welfare effects. All of the estimated coefficients on NTB reduction variables have the expected negative sign, and statistically significant gains of 24,100 US$ result per 1 million US$ increase in blue and green box measures. Reductions in export subsidies, export taxes and amber box measures lead to welfare gains, as does the assumption of exogenous technical change or productivity shocks. The latter effect is highly significant, and indicates that a shock that boosts production by 1 million US$ results in a *ceteris paribus* gain of 79,300 US$.

As expected, the coefficients on the dummy variables for post GTAP-3 databases are negative, increasingly so as one moves from GTAP-4 to GTAP-6. The 'other database' coefficients are uniformly positive and significant, indicating that they generate higher welfare gains than the GTAP databases, *ceteris paribus*, especially when based on applied rather than bound tariffs.

Turning to model characteristics, the estimated coefficients of the 20 dummy variables that depict different combinations of PE/GE, dynamic/static, returns to scale, treatment of capital, and Armington elasticities in GE models display several regularities. The Johansen closure is always associated with larger welfare gains, and high Armington elasticities usually are. High

[6] See footnote 5. Here too, the reference database is used to estimate the value of the measure in question, and this value is subsequently multiplied by the percentage reduction in that measure. Hess and von Cramon-Taubadel (2008) provide details.

Table 1: OLS regression results: Dependent variable is simulated welfare change in million US$

					Weighted regression			Cate
↓Variable					Coefficien	Signif.	Std.	gory
Intercept					-11,639.1	***	1,459.	-
Dependent. variable is absolute change in GDP (2001US$)					4,289.4	***	505.4	MC
Dependent variable is % of baseline EV					4,050.1	***	759.3	MC
Dependent variable is sum of PE surplus and government revenue					-3,441.9	***	981.2	MC
Multi-country GE	Comp. static	CRTS	Capital accum.	Arm. low, 3 primary factors, Johansen	20,519.2	***	3,168.	MC
				Arm. low	313.9		299.4	MC
				Arm. high, 3 primary factors	-3,198.2	***	1,123.	MC
				Arm. high	2,579.4	*	1,469.	MC
		IRTS	Capital fixed	Arm. low	2,723.1	*	1,415.	MC
				Arm. high, Johansen closure	25,612.5	***	5,529.	MC
			Capital accum.	Arm. low	7,458.6	***	723.5	MC
				Arm. high	6,889.4	***	2,186.	MC
	Dyn.c	CRTS	Cap. fix	Arm. high	3,593.0	***	1,241.	MC
			Capital accum.	Arm. low	-2,115.3	*	1,113.	MC
				Arm. high	7,683.7	***	1,068.	MC
		IRTS		Arm. low	2,239.3		2,097.	MC
				Arm. high	13,557.4	***	1,534.	MC
Single country GE	Capital stock fixed			Arm. low	-16,742.8	***	1,299.	MC
	Capital stock accumulation			Arm. low	-1,122.2		687.4	MC
	Capital stock fixed			Arm. high	-19,787.5	***	1,839.	MC
PE	Some primary factors modelled		Shor run	No Arm. assumption	-1,698.6		1,784.	MC
				With Arm. assumption	-1,781.5		1,728.	MC
			Lon run	With Arm. assumption	14,336.4	***	1,364.	MC
	No primary factors			No Arm. assumption	9,177.3	***	2,551.	MC
One or more countries' trade balance fixed					-7,863.8	***	2,845.	MC
Length of dynamic simulation run					-694.0	***	93.3	MC
[Length of dynamic simulation run]²					42.8	***	4.4	MC
Length of pre-simulation projection of database					17.9		51.6	MC
Number of regions depicted					38.0	***	9.1	MC
Number of sectors depicted					402.2	***	14.6	MC
Number of agricultural products depicted					-168.3	***	15.4	MC
Ad hoc modifications to elasticities					3,398.3	***	754.4	MC
Own econometric estimates of elasticities					2,798.4	**	1,245.	MC
Changes in tariff protection					-0.0454		0.029	L
[Changes in tariffs]²					3.5E-08		1.0E-	L
Changes in export subsidies, export taxes and amber box measures					-0.1185		0.086	L
Changes in blue and green box policies					0.0241	**	0.010	L
Changes in non-tariff barriers based on gravity models					-0.0019	***	0.000	L
Changes in non-tariff barriers based on customs documents and other literature					-0.0001		0.000	L
Changes in non-tariff barriers based on observed price wedges (e.g. fob-cif)					-0.0183	***	0.004	L
Shocks to technical change or related variables					0.0793	***	0.007	L
Database GTAP-4					-2,645.3	**	1,324.	DB
Database GTAP-5					-5,782.7	***	936.4	DB
Database GTAP-6					-7,476.5	***	995.0	DB
Non-GTAP database with bound tariffs					6,863.9	***	1,141.	DB
Non-GTAP database with applied tariffs					14,739.4	***	1,578.	DB
Adjusted R²					0.557			-

*Notes: *, ** and *** refer to significance at the 10, 5 and 1 per cent levels, respectively. Standard errors are heteroskedastic consistent.*

Armington elasticities are only associated with negative and significant coefficients in models that do not distinguish between skilled and unskilled labour, and in single country GE specifications, which simulate significantly lower welfare gains when the capital stock is assumed to be fixed. The main effect that is revealed for PE models is that the coefficients corresponding to long run (short run) simulations are always positive (negative). This pattern is statistically significant in the weighted regression, where moving from a short run to a long run PE simulation increases welfare gains by US$14.3 or US$9.2 billion, depending on whether primary factors are modelled and whether the Armington assumption is included in the model. Overall, the results reveal no clear difference between GE and PE applications *per* *se*. Instead, PE and GE models produce larger or smaller simulated welfare gains depending on what other assumptions and features they incorporate (e.g. whether a PE simulation is long run or short run; whether a GE simulation assumes a fixed or an accumulating capital stock, etc.). Long run PE simulations appear to generate larger welfare gains than all but a few GE simulations (in particular, those based on the Johansen closure). This suggests, as a stylised fact, that the dampening impact of small elasticities in PE models tends to outweigh dampening effect of GE linkages in GE models.

Increasing the number of countries/regions and the number of sectors depicted in a model has a positive and significant impact on simulated welfare gains, as expected. The only truly puzzling result in Table 1 is that greater disaggregation of the agricultural sector leads to significantly lower welfare gains. When authors report making modifications to the elasticities in a model, significantly larger welfare gains result. This might indicate that such modifications tend to be undertaken by authors who feel that standard values lead to underestimated welfare gains.

The fixed effect country/region dummies (available from the authors on request) provide some interesting results. The reference country/region is ACP (Africa, Caribbean and Pacific), so the intercept in Table 3 (-11.6 billion US$) can be interpreted as the simulated welfare change for the ACP group of countries that results from the 'average' liberalization scenario and the 'average' model in the literature sample. The fact that this welfare change is negative likely reflects the fact that the ACP countries realize few internal allocation benefits from the average DDA liberalization scenario, in which they are largely exempt from tariff reductions and other policy changes. Furthermore, the ACP countries presumably gain relatively little from increased market access to the rest of the world, and loose as a result of price increases on markets for the major agricultural commodities, for which they are net importers. Against this background, the fixed effect coefficient for Canada amounts to roughly 7.3 billion US$, that for Australia to 5.0 billion US$, and those for the EU-25 and the USA to 14.9 and 16.9 billion US$, respectively (all of these coefficients are significant at the one percent level).

3.4 Preliminary conclusions

The results presented in Hess and von Cramon-Taubadel (2008) and summarized above indicate that a relatively simple meta-regression using variables that describe the liberalisation experiment, the characteristics of the model used, and the database employed can explain an important share of the variation in simulated welfare changes in a sample of DDA trade liberalisation studies. The results provide plausible quantitative estimates of impacts that have hitherto only been considered qualitatively and without accounting for simultaneous variation in numerous factors across modelling frameworks and studies.

Although informative, these results are subject to some limitations. First, many of the independent variables used to estimate equation (1) are dummy variables. Hence, experimentation with more sophisticated functional forms and interaction effects is not possible. Second, many potentially informative studies had to be omitted from the meta-dataset because they do not document even the abbreviated and approximate information that we use to carry out the quantitative analysis presented here.

If documentation were more detailed and complete, more studies could be included in the meta-dataset, and at least some of the qualitative independent variables that we use could be replaced by quantitative alternatives. This would make it possible to use flexible functional forms and consider interactions between independent variables in greater detail. As a result, we expect that a considerably larger proportion of the variation in simulated welfare changes in our sample could be explained. However, while we can hope that the documentation of trade policy simulations will improve in the future, the documentation of past studies is what it is. Furthermore, even if documentation was exhaustive, a limitation of basing a meta-analysis on studies of DDA outcomes is that all of these studies explore similar policy changes and issues. Hence, they tend to focus on only a subset of the space that is spanned by the explanatory variables. Interesting information about the behaviour of trade models might be 'hiding' in other subsets of this space. In the next section we explore the use of meta-analysis based on synthetic data as a means of addressing these limitations.

4.0 META-ANALYSIS OF SYNTHETIC DATA FROM APPLIED TRADE MODELS (RESPONSE SURFACE ANALYSIS)

4.1 Concept and experimental design

Extensive and methodical sensitivity analysis can generate exact, metrically scaled information on many variables that we are only able to extract as dummies from the studies in the DDA literature sample. Furthermore, it can generate observations in subsets of the variable space that are not explored in DDA studies (for example simulations that involve increases rather than reductions in tariffs). The resulting synthetic meta-dataset would make it possible to explore the issues of functional form and interaction between explanatory variables in much greater detail, possibly leading to more powerful insights into the workings of applied trade models.

Multi-dimensional and extensive sensitivity analysis of this nature is not, strictly speaking, meta-analysis, as the latter is a method for combining and evaluating published evidence on research questions. Instead, a more accurate label is 'response surface analysis'. Response surface estimation typically aims to assess the robustness of complex models with many interacting variables. Estimating econometric response surfaces for such models is common in many areas such as engineering, natural sciences and, in economics, especially for agent-based simulations (Kleijnen et al., 2005).

Response surface estimation for a model typically involves an experimental design that generates combinations of the k exogenous model input variables ($X_1, \ldots X_k$) and plugs each combination into the model to simulate a corresponding value of the output (welfare impact) variable (I).[7] This procedure is repeated to generate a dataset (referred to as a 'synthetic meta-dataset' in the following) that is then used to estimate I as a function of ($X_1, \ldots X_k$) econometrically. If a second-order polynomial provides a reasonable approximation, then a suitable econometric response surface model with k factors is a linear model with quadratic and interaction terms (Kutner et al., 2005):

$$E\{I\} = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \beta_{11} X_1^2 + \ldots + \beta_{kk} X_k^2 + \beta_{12} X_1 X_2 + \ldots + \beta_{k-1,k} X_{k-1} X_k \qquad (2)$$

In this model, the coefficients $\beta_1 \ldots \beta_k$ are the linear, $\beta_{11} \ldots \beta_{kk}$ the quadratic and $\beta_{12} \ldots \beta_{k-1,k}$ the interaction term effects. In total, equation (2) requires the estimation of $p=(k+1)(k+2)/2$ parameters. The synthetic meta-dataset for response surface estimation must contain at least three expressions of each variable X to permit estimation of the quadratic terms.

For statistical inference it would be ideal if the synthetic meta-dataset included all possible combinations of the k effects (saturated design). However, for k = 10 the minimum three observations for each factor alone would require a design with 3^{10} = 59,049 combinations of model scenarios to generate the synthetic meta-dataset. It is often argued that computation has

[7] In the notation employed in the literature-based meta-analysis above, the exogenous variables ($X_1, \ldots X_k$) are the variables in the categories MC, LE and DB.

become cheap, but 59,000 simulation runs at two minutes each would require one computer to work for roughly 82 days.

Kutner et al. (2005) as well as Kleijnen et al. (2005) therefore outline practical strategies for less demanding experimental designs. We adopt a Latin hypercube sampling (LHS) strategy. Latin hypercubes are generalizations of Latin square samples. In Latin square samples each combination of factors exists only once. In our context this reduces the computational cost significantly, albeit at the cost of the efficiency of the response surface estimates.

While such sampling strategies can make response surface estimation more manageable, in the case of applied trade models another factor complicates matters. Specifically, the hypothesis that first- and second-order polynomials provide a reasonable approximation for the response surface is questionable as these models are often highly non-linear.[8] While meta regression models typically explain the variance of the dependent variable at an aggregated level for which linear and quadratic approximations are sufficient (Stanley, 2001), meta-modelling of applied trade models should anticipate the potential existence and significance of non-linear model response. As a suitable econometric modelling framework for this purpose, we employ a generalized additive model (GAM) of the following form (Wood, 2006):

$$g(m_i) = \beta_0 + \boldsymbol{\beta}_n \boldsymbol{X}_{ni} + f_1(\boldsymbol{X}_{qi}) + f_2(\boldsymbol{X}_{r-1i}, \boldsymbol{X}_{ri}) + \ldots + \varepsilon_i, \quad (3)$$

where $m_i = \mathrm{E}(I_i)$ and for the application to applied trade models it is assumed that $I_i \sim \mathrm{N}(0, \sigma^2)$, the $\boldsymbol{X}_n$, $\boldsymbol{X}_q$ and $\boldsymbol{X}_r$ are vectors of explanatory variables, and f_1 and f_2 are smooth functions. The number of model input factors to be included in the response surface is $k=n+q+r$. Through specification of the link function g as Gaussian, the parametric parts of the model in the first three terms provide a linear framework that reduces to a generalized linear model (GLM) and, under standard assumptions, is equivalent to the OLS regression model. Note that similar to equation (2), the vector X_n, may also be specified to include interaction effects and/or quadratic terms. The non-parametric parts of the GAM, the functions f_1 and f_2 in equation (3), are estimated using penalized splines (Wood, 2006). The procedure applied for this is penalized iteratively re-weighted least squares (P-IRLS), which we perform using the *mgcv* package of the statistical programming language R. The function f_2 represents a non-parametric interaction term of two explanatory variables.

For response surface modelling of applied trade models, the non-parametric components of equation (3) are important because they facilitate detection and comparison of alternative specifications of functional forms and interaction effects in a unified econometric modelling framework. Similar to meta-regression analysis, the coefficient of determination (adjusted R^2) provides a transparent and well-known criterion for the selection of response surfaces. In addition, an econometric response surface can easily be benchmarked by comparing predicted values against actual simulation results from the trade model in question.

In the following we estimate response surfaces for two applied models of moderate complexity that are calibrated to base data from Canada. In each case a software routine in Visual-Basic is used generate randomly selected combinations of exogenous parameter values chosen from specified ranges (see below). Then, the routine solves the model with these values and saves the model input data and the corresponding output values (simulation results) into a database. The next section describes the models and the specific experimental design that is used.

---

[8] Note, for example, that the systematic sensitivity analysis tool of the standard GTAP model assumes a 3rd degree polynomial approximate model behavior (Arndt, 1996).

4.2 The models used

The main focus of this experiment is to explore methodological aspects of response surface generation for applied trade models, and not to reflect any recent developments in real world trade policy in the base data underlying these models. With this in mind, and to avoid complications that would arise if base data from different sources were merged and compared, GTAP-5 data (base year 1997) is used for both models. Furthermore, to keep the documentation of this response surface experiment transparent and reproducible, *ad hoc* modifications to this data have been avoided.

The Global Simulation Model (GSIM) is a partial equilibrium trade model with Armington-type product differentiation at the regional level. It was developed by Francois and Hall (2003) as a flexible modelling approach that yields insights into trade policy with modest data and parameter requirements. The model covers trade within one industry (sector) between various regions that can be aggregated flexibly. For this paper, an aggregation of four regions has been chosen, but an extended version with 25 regions is also available for download (Francois, 2007). The data required for the GSIM model is limited to bilateral trade flows, bilateral tariffs, supply and demand elasticities as well as substitution elasticities for product differentiation (Armington elasticities) at the country level. The model is available in Excel and can be operated with a spreadsheet solver.

For this paper, the model is calibrated to base data and tariffs for wheat trade between Canada, the US, the EU and the Rest of the World (ROW). Trade flows and bilateral tariffs are obtained from the GTAP-5 dataset. The GTAP-5 tariff data used here do not include preferential agreements and therefore cannot serve as a realistic representation of real world trade policies. For Canada (EU, USA) they show a 62.5 (61, 2.6) percent MFN import tariff on wheat, while ROW regions trade wheat among themselves subject to a 20 percent import tariff and apply 34 percent, 67 percent, and 61 percent tariffs to imports from the EU, Canada and the USA, respectively. Shortcomings of the tariff data in GTAP-5 (and earlier GTAP versions) have received considerable attention in the literature (e.g Bchir et al. 2006). GTAP-6 (base year 2001, released late 2004) uses 'real world' applied tariffs from the MacMap database for the first time; bilateral import tariffs for wheat between Canada and the USA are, as one would expect, zero in these data. In the response surface experiment, however, bilateral tariffs in the GSIM model have been allowed to vary simultaneously and therefore effectively represent preferential trade agreements around each region's MFN mean.

To generate the synthetic meta-data, all elasticities are varied between |0.01| and |5|. Trade flows are varied between 0 and 20 billion US$, and tariff changes for each generated scenario are varied between 0 to 100 percent of the original bilateral GTAP-5 tariffs. This implies that simulated ordinary tariff changes for Canada are in the range of -80 percent to +60 percent for imported wheat, depending on the initial bilateral tariff level that has been obtained from the GTAP-5 database.

Since each simulation run under this regional aggregation only requires several seconds, we decided to also generate the synthetic meta-data using two alternative solver specifications: First, simulations are run based on the default solver settings that are specified in the downloaded model. Second, the same simulations are conducted using a more conservative solver specification (1,000 iterations; accuracy 10^{-6}, tolerance 10^{-5}%; convergence 10^{-8}; automatic scaling; cubic estimates; Newton method). For each solver specification, the model is solved for 10,000 combinations of exogenous variable values.

The second model used for response surface generation is a single country CGE model benchmarked to GTAP-5 pre-release data for Canada. The model was developed by van der Mensbrugghe (2000) to facilitate flexible trade policy analysis in a general equilibrium framework. The model contains many features of a typical single country CGE, where 'typical' refers to the average values found in the literature meta-dataset employed in section 3. These features include: production nests based on CES functions distinguish primary and intermediate factors originating from domestic and imported sources; Armington product differentiation for domestic and imported products on the demand side; private consumption based on an extended linear expenditure system; a neoclassical-style macro-closure with savings-driven investment; and government revenues equal to government spending that is determined by direct and indirect taxes on domestic goods, imports and domestic consumption.

Canada's trade balance is fixed in this model which implies that trade liberalization does not trigger increased inflows of foreign capital, which in turn is in fixed supply. The numeraire of the model is the 'real' exchange rate specified as an index of foreign prices against which domestic prices rise or fall in relative terms as a result of policy changes. The model covers only two aggregated sectors, all agricultural (AGR) and all other non-agricultural (OTH) products produced and consumed in Canada.

In the experimental design, tariffs and export subsidies/taxes are allowed to vary +/-100 percent around the default GTAP tariffs. Again, trade parameters for input substitution, export substitution and export demand in AGR and OTH, respectively, are allowed to vary between 0.01 and up to 5 times their original values. If the VBA routine happens to set any parameter values to one, the scenario is eliminated from the dataset, since this value induces a breakdown in the convergence process (van der Mensbrugghe, 2000).

Finally, variability is also introduced into the model's social account matrix (SAM). Thorbecke (2001) suggests that due to measurement and aggregation error it would be more convincing to consider the SAM as a stochastic rather than as a deterministic depiction of the input-output relationships in an economy. It is not clear exactly how uncertainty in SAMs should be incorporated in response surface estimation, but as a first attempt the following procedure is applied to the SAM for Canada in the single country CGE model. For each observation in the synthetic meta-dataset, roughly one third of the SAM entries are allowed to vary within a range of +/- 50 percent about their original values (obtained from the GTAP dataset). Table 2 presents the original SAM and highlights the entries that have been allowed to vary. An iterative Visual Basic routine then adjusts the remaining SAM entries to ensure that the accounting restriction $\Sigma_{\text{rows}} - \Sigma_{\text{columns}} = 0$ is maintained. While somewhat *ad hoc*, this procedure nevertheless makes it possible to estimate the sensitivity of the CGE model with respect to moderate changes in the base data composition. Such changes could be the result of yearly fluctuations in prices, trade flows, etc., or of inaccuracies and measurement error that are introduced when data for real SAMs are assembled (Thorbecke, 2001). However, we admit that this approach only maintains the condition that a SAM is balanced overall; it does not ensure that various sub-relationships within the SAM necessarily hold (compare e.g. Reinert and Roland-Holst, 1997).

Solving the single country CGE is more time consuming than the GSIM model; therefore, a smaller sample of 1,000 scenarios is solved under default solver settings, and for the more conservative solver settings described above the model is solved about 10,000 times.

Table 2: Social account matrix for Canada. Entries that are allowed to vary in bold, estimated coefficients (see Table 5) in parentheses

	AGR	OTH	Labour	Capital	Other (land)	Income tax	Export tax	House-hold	Govern ment	Invest-ment	De-precia-tion	Market margin	Tariff revenue	Rest of world	Sum
AGR	**7924** (-0.031)	**23188** (0.015)					38	4930	1095	3		0		**10154** (0.011)	47333
OTH	**18736** (-0.007)	**464809** (0.001)					459	330726	122835	114982		3238		**225080** (-0.007)	1280863
Labour	**8424** (-0.024)	**312781** (0.004)													321205
Capital	**6977** (-0.038)	**213616** (0.003)													220594
Other	**3531** (0.010)	**5316** (-0.052)													8848
Income tax	**-2105** (0.101)	**12695** (-0.000)						34750	517	3303					49160
Export tax									497						497
House-holds			321205	220594	8848				0						550647
Govern -ment	**-691** (-0.274)	**36215** (0.002)	0			49160		36765					3495		124943
Invest-ment								80430	0		63046				143476
Deprec.								63046							63046
Market margin														3238	3238
Tariff revenue	**23** (-0.068)	**3471** (0.062)													3495
Rest of world	4513	208771								25187					238472
Sum	47333	1280863	321205	220594	8848	49160	497	550647	124943	143476	63046	3238	3495	238472	3055814

Source: van der Mensbrugghe (2000), coefficients from own estimates.

It would be possible to employing more up-to-date data (GTAP-6) in these experiments, but this would constitute a time consuming task especially in the single country CGE model. This task has not yet been undertaken, partly because the GTAP-7 data base (base year 2004) is currently under development and will soon replace GTAP-6. While the use of different base data could be expected to change the estimated coefficients of the response surface, it would be unlikely to affect our general methodological insights regarding the quantitative comparisons of applied trade models.

4.3 Results

Regardless of the solver settings, the GSIM model converges for all scenarios. The single country CGE fails to converge for about 0.3 percent of the simulation runs. This share is not significantly affected by solver specifications; a probit regression on the non-converging scenarios suggests that the trade elasticities are correlated with failed convergence, implying that specific combinations of these elasticities rather than overall solver settings lead to failed convergence. Scenarios that did not converge are eliminated from the synthetic meta-dataset.
The coefficient of determination of the estimated response surface is, depending on model specification, about 10 percent to 15 percent lower when less restrictive solver settings are used. This result may not surprise experienced modellers and is perhaps irrelevant for models based on GEMPACK (GTAP) or GAMs, where automatic accuracy checks can be performed during the solving procedure. However, this result suggests that even if all functional forms and interaction terms for the two models were known with certainty, it would nevertheless be impossible to obtain a perfect fit for the corresponding response surfaces, even though the models are deterministic. Solver inaccuracies appear to introduce non-negligible noise into numerical results.

For the PE as well as for the CGE model, equation (3) is first estimated with only first order terms and without interaction effects; due to the Gaussian link function it is therefore identical to an OLS linear regression. This regression produces an adjusted R^2 for the PE (CGE) model of 77 percent (49 percent). Starting from these response surfaces, all exogenous parameters in each model are next estimated using penalized tensored splines to detect higher order functional forms. The base- and smoothing parameters of these penalized splines are specified according to standard assumptions in order to avoid over-fitting (see R *mgcv* package by Wood, 2006).

If only polynomial forms of the model response are modelled this way, but interaction effects are ignored, the adjusted R^2 only increases from 77 to 78.9 percent for GSIM; for the CGE it increases from 49 to 55 percent. In both models, up to fourth-order polynomials are detected. Alternatively, if interaction terms between independent variables are added to the model, the adjusted R^2 increases to 87 percent for GSIM and to 80 percent for the CGE. Furthermore, if all non-parametric splines are removed from the model and instead squared terms for tariffs are included along with the most significant interaction effects, the adjusted R^2 does not drop for GSIM (Table 3). In the case of the CGE, squared terms for tariffs, and those elasticities for which higher order polynomials were indicated, are included in the response surface equation. Altogether, eight variables are included: import tariffs (AGR, OTH), export taxes (AGR, OTH), import- and export substitution elasticities (AGR, OTH) and transformation elasticities between import and export supply (AGR, OTH). To keep the regression model parsimonious only the most significant interaction effects are retained. Consequently, the adjusted R^2 drops from 80 to 70 percent (Table 4).

The resulting response surface models in Tables 3 and 4 provide first- and second-order approximations to each model response $E\{Y_i\}$ while retaining a major advantage of parametric

Table 3: Response surface estimates for GSIM (partial equilibrium model)

Estimate	Coefficien	Std.	t-value	Pr(>|t|)	Signif.
Intercept	148.163	183.304	0.808	0.42	
Tradevolume USACanada * ΔTariff	0.075	0.007	10.457	0.00	***
Tradevolume USAEU * ΔTariff	0.114	0.012	9.557	0.00	***
Tradevolume USAROW * ΔTariff	-1.022	0.012	-87.275	0.00	***
Tradevolume CanadaUSA * ΔTariff	0.165	0.018	9.313	0.00	***
Tradevolume CanadaEU * ΔTariff	0.109	0.012	9.343	0.00	***
Tradevolume CanadaROW * ΔTariff	-0.954	0.015	-65.574	0.00	***
Tradevolume EUUSA * ΔTariff	0.184	0.018	10.366	0.00	***
Tradevolume EUCanada * ΔTariff	0.085	0.009	9.065	0.00	***
Tradevolume EUROW * ΔTariff	-1.021	0.012	-84.955	0.00	***
Tradevolume ROWUSA * ΔTariff	0.224	0.018	12.456	0.00	***
Tradevolume ROWCanada * ΔTariff	0.071	0.009	7.632	0.00	***
Tradevolume ROWEU * ΔTariff	0.100	0.012	8.346	0.00	***
Tradevolume ROWROW * ΔTariff	-1.011	0.014	-73.817	0.00	***
Tradevolume EUEU	-0.021	0.005	-4.570	0.00	***
(Tradevolume USAEU * ΔTariff)^2	0.00001	0.00000	2.289	0.02	*
(Tradevolume USAROW * ΔTariff)^2	0.00001	0.00000	3.677	0.00	***
(Tradevolume CanadaUSA * ΔTariff)^2	0.00000	0.00000	2.068	0.04	*
(Tradevolume CanadaEU * ΔTar.CA-EU)^2	0.00000	0.00000	-0.482	0.63	
(Tradevolume CanadaROW * ΔTariff)^2	0.00002	0.00000	7.667	0.00	***
(Tradevolume EUUSA * ΔTariff)^2	-0.00001	0.00000	-4.098	0.04	***
(Tradevolume EUCanada * ΔTariff)^2	0.00000	0.00000	-1.623	0.10	
(Tradevolume EUROW * ΔTariff)^2	0.00002	0.00000	8.497	0.00	***
(Tradevolume ROWUSA * ΔTariff)^2	-0.00001	0.00000	-5.291	0.00	***
(Tradevolume ROWCanada * ΔTariff)^2	0.00000	0.00000	-2.834	0.00	**
(Tradevolume ROWEU * ΔTariff)^2	-0.00001	0.00000	-3.752	0.00	***
(Tradevolume ROWROW * ΔTariff)^2	0.00001	0.00000	5.948	0.00	***
Demand elasticity USA	-810.961	17.300	-46.876	0.00	***
Supply elasticity USA	3.619	18.012	0.201	0.84	
Substitution elasticity USA	-12.069	17.915	-0.674	0.50	
Demand elasticity Canada	147.739	17.410	8.486	0.00	***
Supply elasticity Canada	-67.984	17.285	-3.933	0.00	***
Substitution elasticity Canada	-52.583	17.490	-3.006	0.00	**
Demand elasticity EU	277.268	17.502	15.842	0.00	***
Supply elasticity EU	-162.872	17.522	-9.296	0.00	***
Substitution elasticity EU	-40.369	18.151	-2.224	0.03	*
Demand elasticity ROW	43.436	18.936	2.294	0.02	*
Supply elasticity. ROW	-223.931	18.863	-11.871	0.00	***
Substitution elasticity ROW	93.536	18.072	5.176	0.00	***
Residual standard error: 2460		F-statistic: 1789 on 38 and 10011			
Multiple R-Squared: 0.87		Adjusted R-squared: 0.87			

*Note: *, ** and *** refer to significance at the 10, 5 and 1 per cent levels, respectively.*

Table 4: Single country CGE response surface for Canada

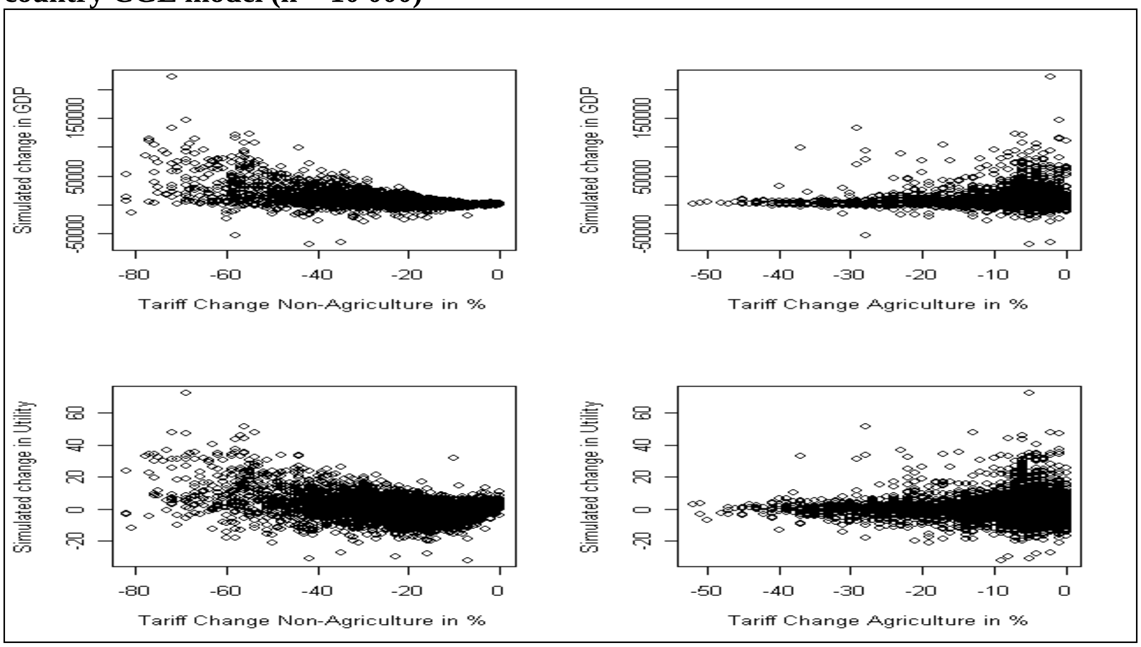
Variable	Coefficient	Std. error	t-value	Pr(>|t|)	Signif.
Intercept	-4025.8	807.8	-4.98	0.000	***
Δ Tariff Other Sectors (Oth)	44411.48	2783.8	15.95	0.000	***
Δ Tariff Agriculture (Agr)	-6766.8	5668.1	-1.19	0.233	
'Armington' CES parameter for substitution imports/domestic products 'Other Sectors': σ_m_Oth	-71.18	38.0	-1.87	0.061	
CET parameter exports/domestic production 'Other Sectors': σ_x_Oth	91.38	36.2	2.52	0.012	*
CET parameter exports/domestic production Agriculture: σ_x_Agr	9.1	35.0	0.26	0.795	
'Armington' CES parameter for substitution imports/domestic products Agriculture: σ_m_Agr	-8.5	35.0	-0.24	0.809	
Δ Export Tax Agr	-4288.1	3982.6	-1.08	0.282	
Δ Export Tax Oth	-32028.6	8138.3	-3.94	0.000	***
$(\Delta \text{ Export Tax Agr})^2$	40306.9	19511.3	2.07	0.039	*
$(\Delta \text{ Export Tax Oth})^2$	-251701.4	84071.1	-2.99	0.003	**
$(\Delta \text{ Tariff Agr})^2$	2107.2	9782.1	0.22	0.830	
$(\Delta \text{ Tariff Oth})^2$	101486.7	2994.6	33.89	0.000	***
Elasticity of foreign export demand Agr	83.9	56.1	1.50	0.135	
Elasticity of foreign export demand Oth	1465.7	56.1	26.13	0.000	***
$(\text{Elasticity of foreign export demand Agr})^2$	-6.8	3.3	-2.07	0.039	*
$(\text{Elasticity of foreign export demand Oth})^2$	-69.8	3.3	-21.09	0.000	***
Δ Tariff Oth * σ_m_Oth	-1285.7	188.2	-6.83	0.000	***
Δ Tariff Oth * σ_m_Oth	701.8	178.0	3.94	0.000	***
σ_m_Oth * σ_x_Oth	-40.8	4.3	-9.49	0.000	***
σ_x_Agr * σ_m_Agr	-0.7	4.0	-0.17	0.869	
σ_x_Agr * Δ Tariff Agr	995.7	362.1	2.75	0.006	**
σ_m_Agr * Δ Tariff Agr	661.0	351.8	1.88	0.060	.
Δ Export Tax Agr * Δ Export Tax Oth	-83283.5	81587.4	-1.02	0.307	
El.'s foreign exp dem. (Agr * Oth)	7.6	3.0	2.53	0.012	*
ΔTariff Oth * σ_m_Agr * σ_x_Oth	-526.4	21.1	-24.96	0.000	***
ΔTariff Agr * σ_x_Agr * σ_m_Agr	-101.2	40.0	-2.53	0.011	*
Residual standard error = 5570					
Multiple $R^2 = 0.7018$, adjusted $R^2 = 0.7005$					
F-statistic: 559.8 on 44 and 10468					

*Note: *, ** and *** refer to significance at the 10, 5 and 1 per cent levels, respectively.*

regression (i.e. convenient interpretation of marginal effects). In the following we highlight several key results of the response surface estimations:

- Both sets of results confirm that simulated GDP and welfare effects can vary widely for the same tariff reduction experiment depending on the values of other parameters in the model. This is illustrated in Figure 1, which plots for the CGE simulated GDP / utility changes against tariff reductions for the single country CGE model. Figure 1 shows that the variance of the simulated welfare gains is large and varies with the size of the liberalization step if the impact of other covariates and interactions terms are not controlled for.

Figure 1: Simulated GDP/ utility changes as a function of tariff changes in the single country CGE model (n = 10 000)



- The high R^2 values in Tables 3 and 4 (in the range of 70-80 percent) confirm that exact knowledge of all model characteristics and other factors that go into a trade policy simulation makes it possible to explain a much higher proportion of this variance in simulation outcomes than is possible given only the information that can be extracted from publications that report the results of such simulations. Recall that the R^2 reported for the literature based meta-regression in section 3 was roughly 57 percent. This R^2 cannot be compared with those here, however, because it is based on a weighted regression (i.e. a transformed independent variable). Comparable R^2 values from unweighted regressions reported in Hess and von Cramon-Taubadel (2008) range from 24 to 47 percent. Hence, the availability of detailed information on exactly what went into a model simulation (as opposed to the incomplete and approximate information that we were able to extract from the literature sample) leads to a large improvement in the explanatory power of the estimated meta-regression.

- The results in Table 3 confirm that GSIM results are fundamentally driven by initial tariff levels and trade volumes. Welfare effects based on calculations of economic rents are related to the square of the tariff change. With regard to the magnitude and sign of the coefficients it is not immediately clear why Canada only experiences a net welfare gain when it or any other country reduces its tariffs vis-à-vis ROW, while tariff cuts vis-à-vis any other trade partner are net welfare decreasing. The reason is that in this experimental setting, high and low initial tariffs are not altered proportionally. Instead, all tariffs are subject to the same range of random changes. Whenever initial bilateral tariffs are below the average new tariff, this results in an average tariff increase in the synthetic meta-data, while only countries that initially show tariffs higher than average experience an average net tariff cut. Furthermore, since average trade volumes and elasticities are equal for all countries in the synthetic meta-data, price changes in this model and thus the estimated coefficients in Table 3 are directly related to the initial bilateral tariffs in absolute terms as well relative to each other.

- Table 4 presents the response surface for the single country CGE model, with the exception of the estimated coefficients for the SAM base data which are presented in Table 5 (and can also be seen in Table 2). With regard to the magnitude of estimated coefficients Table 4 shows that changes in variables related to AGR in general have a much smaller effect on GDP in Canada. This indicates – as one would expect – that GE effects even from an aggregated agricultural sector play a minor role in the overall Canadian economy.

- The estimated coefficients for SAM base data in Table 5 indicate by how much Canada's real GDP in million US$ changes if the corresponding SAM entries change by US$1 million and the rest of the SAM is adjusted so as to maintain the accounting identity $\Sigma_{rows} - \Sigma_{columns} = 0$. Interestingly, the largest coefficients are for government expenditure on agriculture and income tax revenues from agriculture. The former (-0.274) indicates that a US$1 million reduction in government spending would increase real GDP by US$274,000, while the latter (0.101) indicates that reducing tax expenditures and transfer payments to agriculture would increase real GDP by US$101,000. These effects are not or only just significant at conventional levels, but are suggestive of the especially distorting impact of agricultural policies and the economic burden that agriculture places on the economy as a whole. With regard to the statistical significance levels it should be emphasized that due to the synthetic structure of the meta-dataset (e.g. the arbitrarily chosen number of simulation runs), a conventional interpretation of t-values is not possible. Furthermore, within this reasoning no effort has (yet) been made to apply robust standard errors to the estimated coefficients.

- In both models, interaction effects between key input parameters account for much more of the variance in the dependent variable than higher-order (greater than 2) polynomial effects. Moreover, this impact of interaction effects is much stronger in the CGE model than in the PE model. This confirms one of the results of the literature-based meta-analysis in section 3, in which numerous interaction effects (captured relatively crudely in extensive sets of dummy variables for specific combinations of model characteristics) are found to have significant explanatory power.

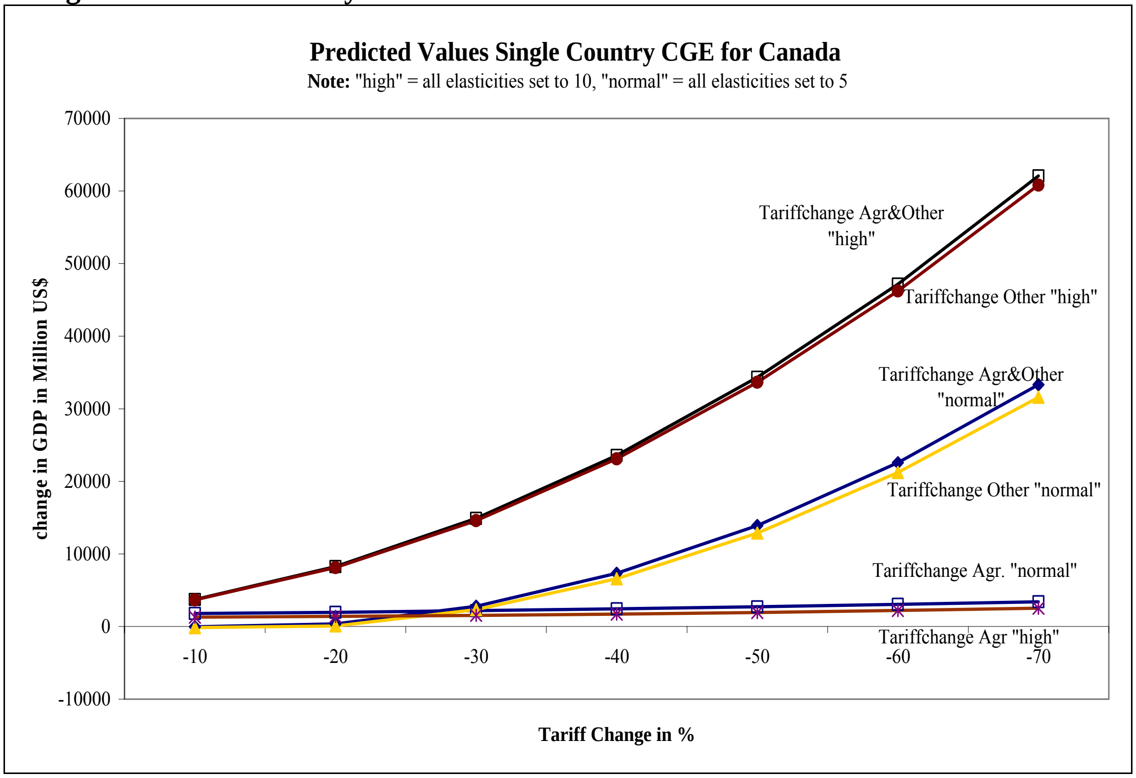
- Figure 2 illustrates the relationship between tariff changes and predicted GDP changes from the response surface results in Table 4 under different assumptions regarding the size of the elasticities in the model (utilizing sample averages for all other variables except the tariff cuts displayed in Figure 2). As one would expect, the effect of tariff cuts in agriculture and/or in the other sectors is strictly positive for Canada's GDP. When other elasticities are high, higher welfare effects result from a given tariff change than when the other elasticities are low. However, reducing agricultural tariffs alone affects GDP only moderately compared with tariff reductions in the rest of the economy. When both agricultural and other tariffs are reduced simultaneously, the predicted change in GDP is

Table 5: Estimated coefficients for the SAM base data, single country CGE for Canada.

Variable	Coefficient	Std. error	t-value	Pr(>|t|)	Signif.
AgrAgr	-0.0317	0.0149	-2.1220	0.0339	**
AgrOth	0.0157	0.0067	2.3350	0.0196	**
AgrLab	-0.0241	0.0144	-1.6740	0.0942	
AgrKap	-0.0380	0.0171	-2.2180	0.0266	**
AgrFct	0.0100	0.0335	0.2990	0.7646	
AgrItax	0.1012	0.0563	1.7990	0.0720	*
AgrGov	-0.2749	0.1712	-1.6060	0.1083	
AgrTar	-0.0689	0.9346	-0.0740	0.9413	
OthAgr	-0.0073	0.0056	-1.3120	0.1897	
OthOth	0.0011	0.0003	4.1970	0.0000	***
OthLab	0.0042	0.0004	10.4780	0.0000	***
OthKap	0.0031	0.0006	5.4450	0.0000	***
OthFct	-0.0527	0.0223	-2.3630	0.0181	**
OthItax	-0.0004	0.0093	-0.0460	0.9631	
OthGov	0.0025	0.0033	0.7740	0.4390	
OthTar	0.0624	0.0120	5.2170	0.0000	***
OthRow	-0.0068	0.0012	-5.6410	0.0000	***
AgrRow	0.0113	0.0525	0.2160	0.8290	

*Note: *, ** and *** refer to significance at the 10, 5 and 1 per cent levels, respectively.*
The abbreviations refer to Table 4 and label SAM entries from 'row' to 'column'.

Figure 2: Predicted values for the single country CGE for Canada as a function of tariff changes and other elasticity values



Note: High = all elasticities set to 10; normal = all elasticities set to 5.

almost identical to that when tariffs in OTH are reduced alone. Finally, the effect of simultaneous changes in both agricultural and other tariffs are smaller than the sum of the effects of individual reductions in agricultural and other tariffs.

So far, both trade models have been analyzed separately, and the estimated response surfaces provide more or less the same insights that a thorough and comprehensive sensitivity analysis could. In the spirit of meta-analysis we next attempt to estimate a joint response surface for both models. The dependent variable in GSIM is the change in consumer and producer rents (the GSIM model does not incorporate tax revenue into the welfare measure), while the single country CGE is solved for changes in GDP. To merge the individual synthetic meta-datasets from these models, the simulated change in consumer surplus is taken from the GSIM results, and simulated change in consumer utility is taken from the single country CGE[9]. The result is a new dependent variable labelled 'Welfarechange'. Note that the question which dependent variables from the respective models are most appropriate for comparison does not affect the feasibility of the response surface for both models. To control for differences between the measures, a dummy variable (PE=1 if the observation in questions stems from GSIM) is included in the regression. For all explanatory variables that are included in one model but not the other, missing values are imputed using sample means (Greene, 2003; Little, 1992).

Table 6 presents estimation results for this combined synthetic meta-dataset, where the coefficient of determination as well as the signs, magnitudes and significance levels of most explanatory variables are similar to those in Tables 3 and 4. The coefficient of the PE dummy shows that after controlling for all other effects that are captured by the explanatory variables, the measure of consumer surplus from the GSIM model is *ceteris paribus* US$4.7 billion higher (with a standard error of US$116 million) than the change in consumer utility in the CGE model. This preliminary result indicates that the joint estimation of one response surface for both models is feasible and able to generate econometric measures of the difference between simulation output from two very distinct applied trade models.

5. CONCLUSION

Neither conventional sensitivity analysis nor qualitative reviews of simulation models and their results can provide satisfying explanations for the sometimes very large differences between trade simulation results. In this paper we study whether meta-analysis can provide such explanations.

We first summarise the results of a meta-analysis (documented thoroughly in Hess and von Cramon-Taubadel, 2008) based on 110 published Doha Development Agenda simulations that produces plausible and robust results. However, the sample that can be drawn from the literature suffers from several limitations. In particular, poor documentation in many cases leads to sample attrition and often only stylized or approximate (i.e. dummy variable) depictions of key model characteristics.

[9] Alternatively, one might compare the entire welfare measure from GSIM to changes in Utility from the CGE model. For a discussion of related welfare measures see Mas-Colell et al. 1995.

Table 6: Changes in consumer utility (CGE) and consumer surplus (PE) jointly estimated using synthetic data from both models

Estimate	Coefficient	Std. error	t-value	Pr(>|t|)	Significance
Intercept	-12511.08	513.268	-24.375	0.000	***
Tradevolume USACanada * ΔTariff	-1.11	0.009	-121.083	0.000	***
Tradevolume USAEU * ΔTariff	0.13	0.015	8.440	0.000	***
Tradevolume USAROW * ΔTariff	0.10	0.015	6.733	0.000	***
Tradevolume CanadaUSA * ΔTariff	-0.06	0.023	-2.612	0.009	**
Tradevolume CanadaEU * ΔTariff	-0.00	0.015	-0.129	0.897	
Tradevolume CanadaROW * ΔTariff	-0.03	0.019	-1.673	0.094	.
Tradevolume EUUSA * ΔTariff	0.17	0.023	7.669	0.000	***
Tradevolume EUCanada * ΔTariff	-0.95	0.012	-78.613	0.000	***
Tradevolume EUROW * ΔTariff	0.11	0.015	7.227	0.000	***
Tradevolume ROWUSA * ΔTariff	0.18	0.023	7.664	0.000	***
Tradevolume ROWCanada * ΔTariff	-0.95	0.012	-79.609	0.000	***
Tradevolume ROWEU * ΔTariff	0.06	0.015	4.180	0.000	***
Tradevolume ROWROW * ΔTariff	0.10	0.018	5.797	0.000	***
Tradevolume EUEU	-0.02	0.006	-3.443	0.001	***
(Tradevolume USAEU * ΔTariff)^2	0.0000	0.0000	1.471	0.141	
(Tradevolume USAROW * ΔTariff)^2	0.0000	0.0000	-0.288	0.773	
(Tradevolume CanadaUSA * ΔTariff)^2	0.0000	0.0000	3.917	0.000	***
(Tradevolume CanadaEU DelTarCanadaEU)^2	0.0000	0.0000	0.722	0.471	
(Tradevolume CanadaROW * ΔTariff)^2	0.0000	0.0000	-2.061	0.039	*
(Tradevolume EUUSA * ΔTariff)^2	0.0000	0.0000	-2.142	0.032	*
(Tradevolume EUCanada * ΔTariff)^2	0.0000	0.0000	15.276	0.000	***
(Tradevolume EUROW * ΔTariff)^2	0.0000	0.0000	1.489	0.136	
(Tradevolume ROWUSA * ΔTariff)^2	0.0000	0.0000	-1.555	0.120	
(Tradevolume ROWCanada * ΔTariff)^2	0.0000	0.0000	12.733	0.000	***
(Tradevolume ROWEU * ΔTariff)^2	0.0000	0.0000	-3.737	0.000	***
(Tradevolume ROWROW * ΔTariff)^2	0.0000	0.0000	1.533	0.125	
Demand elasticity USA	-508.67	22.229	-22.883	0.000	***
Supply elasticity USA	55.39	23.144	2.393	0.017	*
Substitution elasticity USA	-142.03	22.371	-6.349	0.000	***
Demand elasticity Canada	-21.98	22.210	-0.990	0.322	
Supply elasticity Canada	114.75	22.473	5.106	0.000	***
Substitution elasticity Canada	1.91	23.020	0.083	0.934	
Demand elasticity EU	-29.46	23.323	-1.263	0.207	
Supply elasticity EU	25.05	23.222	1.079	0.281	
Substitution elasticity EU	244.00	22.488	10.850	0.000	***
Demand elasticity ROW	-162.09	22.514	-7.200	0.000	***
Supply elasticity ROW	75.88	24.332	3.119	0.002	**
Substitution elasticity ROW	-192.94	24.238	-7.960	0.000	***
Δ Tariff Other Sectors ('Oth')	-49.53	15.789	-3.137	0.002	**
Δ Tariff Agricultural Sector ('Agr')	-57.33	21.975	-2.609	0.009	**
'Armington' CES parameter import/dom: σ_m_Oth	-54.06	21.539	-2.510	0.012	*
CET parameter export/dom. production: σ_x_Oth	8.80	20.552	0.428	0.669	
CET parameter export/dom :production σ_x_Agr	31.38	19.857	1.580	0.114	
'Armington'. CES parameter import/dom: σ_m_Agr	30.46	19.859	1.534	0.125	
Δ Export Tax Agr	1687.53	2259.701	0.747	0.455	
Δ Export Tax Oth	59529.20	4617.703	12.892	0.000	***

(Δ Export Tax Agr)^2		-5509.39	11070.629	-0.498	0.619	
(Δ Export Tax Oth)^2		-708176.13	47702.947	-14.846	0.000	***
(Δ Tariff Agr)^2		-0.55	0.368	-1.499	0.134	
(Δ Tariff Oth)^2		2.90	0.170	17.043	0.000	***
Elasticity of foreign export demand Agr		-44.82	31.823	-1.409	0.159	
Elasticity of foreign export demand Oth		1252.71	31.830	39.357	0.000	***
(Elasticity of foreign export demand Agr)^2		4.15	1.880	2.209	0.027	*
(Elasticity of foreign export demand Oth)^2		-52.96	1.879	-28.192	0.000	***
SAM coefficients, please compare Tables 2 and Table 5.	AgrAgr	0.0032	0.009	0.380	0.704	
	AgrOth	0.0091	0.004	2.401	0.016	*
	AgrLab	-0.0057	0.008	-0.701	0.483	
	AgrKap	-0.0013	0.010	-0.134	0.893	
	AgrFct	0.0248	0.019	1.303	0.193	
	AgrItax	0.0205	0.032	0.644	0.520	
	AgrGov	-0.0022	0.097	-0.023	0.982	
	OthAgr	-0.0125	0.003	-3.952	0.000	***
	OthOth	-0.0011	0.000	-7.440	0.000	***
	OthLab	0.0076	0.000	33.502	0.000	***
	OthKap	0.0069	0.000	21.416	0.000	***
	OthFct	-0.0141	0.013	-1.113	0.266	
	OthItax	0.0010	0.005	0.197	0.844	
	OthGov	0.0091	0.002	4.895	0.000	***
	OthTar	-0.2122	0.007	-31.282	0.000	***
	OthRow	-0.0028	0.001	-4.180	0.000	***
	AgrRow	-0.0455	0.018	-2.472	0.014	*
PE dummy (1 if GSIM partial equilibrium model)		4757.43	116.979	40.669	0.000	***
Δ Tariff Oth * σ_m_Oth		-11.20	1.068	-10.489	0.000	***
Δ Tariff Oth * σ_m_Oth		10.09	1.010	9.990	0.000	***
σ_m_Oth * σ_x_Oth		-12.15	2.437	-4.988	0.000	***
σ_x_Agr * σ_m_Agr		-3.49	2.276	-1.531	0.126	
σ_x_Agr * Δ Tariff Agr		6.32	2.053	3.077	0.002	**
σ_m_Agr * Δ Tariff Agr		4.90	1.992	2.458	0.014	*
Δ Export Tax Agr * Δ Export Tax Oth		68035.62	46290.516	1.470	0.142	
El.'s foreign exp dem. (Agr * Oth)		-1.64	1.712	-0.955	0.340	
ΔTariff Oth * σ_m_Agr * σ_x_Oth		-0.49	0.120	-4.078	0.000	***
ΔTariff Agr * σ_x_Agr * σ_m_Agr		-0.75	0.227	-3.313	0.001	***
Residual sum of squares = 3161						
F-statistic = 873.2 on 82 and 20480						
Multiple R² = 0.7776; adjusted R² = 0.7767						

*Note: *, ** and *** refer to significance at the 10, 5 and 1 per cent levels, respectively.*

Second, we present the results of a meta-analysis using synthetic data (response surface analysis) that can mitigate these limitations. If too few well-documented published studies are available for traditional meta-analysis, the estimation of response surface enables both direct comparison of output and input from different models, and detailed quantitative assessment of the impact of individual modelling frameworks, parameters and base data specifications. The response surface analysis presented here is preliminary and can clearly be refined and better tailored to specific tasks. Issues that require more detailed analysis include the influence of solvers and solver settings on simulation output including the state-of-the-art of modern solver techniques used in multi-region PE and CGE models, and the use of more sophisticated non-parametric estimation techniques. The trade-off between the complexity of a response surface and ease of interpretation should be kept in mind when pursuing the latter. A related question is whether it is possible to develop response surfaces that could offer a low-cost alternative to modelling, at least to a first degree of approximation.

By adding to our understanding of how model characteristics, liberalisation experiments and databases influence trade policy simulations, meta-analysis can contribute to reducing the impression of arbitrariness that arises when economists produce what appear to be very different estimates of liberalization benefits. The quantitative estimates of individual impacts reported here can be used by both modellers and model users to compare and at least partially reconcile divergent simulation results. Exercises of this nature can be especially beneficial for low income countries, which often cannot afford to maintain sophisticated own modelling capacities and dedicate highly trained personnel to the comparison and assessment of different and often conflicting model results.

REFERENCES

Anderson, K., Martin, W. and van der Mensbrugghe, D., 2006. Doha Merchandise Trade Reform: What's at Stake for Developing Countries? *World Bank Economic Review*, 20(2): 169-195.

Arndt, C., 1996. *An Introduction to Systematic Sensitivity Analysis via Gaussian Quadrature*. GTAP Technical Paper No. 2. https://www.gtap.agecon.purdue.edu/resources/ res_display.asp?RecordID=305.

Bchir, M. H., Jean, S. and Laborde, D., 2006. Binding Overhang and Tariff-Cutting Formulas. In: *Review of World Economics*, 142(2): 207-232.

Boys, K. A. and Florax, R. J. G. M., 2007. *Meta-Regression Estimates for CGE Models: A Case Study for Input Substitution Elasticities in Production Agriculture*. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29 – August 1, 2007.

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

Francois, J., 2007. *Applied Trade Models for Download*. Internet-Resource. http://www.intereconomics.com/handbook/Models/Index.htm.

Francois, J. and Hall, K., 2003. *Global Simulation Analysis of Industry-Level Trade Policy*. Version 3. Working paper: http://www.intereconomics.com/handbook/Models/Index.htm.

Greene, W., 2003. *Econometric Analysis*. Prentice-Hall, 5th Edition, Upper Saddle River, USA.

Hertel, T. W., 1999. *Future Directions in Global Trade Analysis*. Staff Paper #99-8, Department of Agricultural Economics, Purdue University.

Hess, S. und von Cramon-Taubadel, S., 2008. Doha Round Outcomes: A Meta-Analysis of General and Partial Equilibrium Simulations. Forthcoming in *The World Economy,* Vol. 31(6): 804-840.

Kleijnen, J. P. C., Sanchez, S., Luca, T. and Cioppa, T., 2005. A User's Guide to the Brave New World of Designing Simulation Experiments. *INFORMS Journal on Computing*, 17(3): 263-289.

Knell, M. and Stix, H., 2005. The Income Elasticity of Money Demand: A Meta-Analysis of Empirical Results. *Journal of Economic Surveys*, 19(3): 513-533.

Kutner, M., Nachtsheim, C., Neter, J. and Li, W., 2005. *Applied Linear Statistical Models*. 5th Edition McGraw-Hill, New York.

Little, R., 1992. Regression With Missing X's: A Review. *Journal of the American Statistical Association*, 87(420): 1227-1237.

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

Piermartini, R. and Teh, R., 2005. *Demystifying Modelling Methods for Trade Policy*. Discussion Paper No. 10, WTO (World Trade Organization), Geneva.

Reinert, K. and Roland-Holst, D., 1997. Social Account Matrices. In: Francois, J. and Reinert, K., (eds.): *Applied Methods for Trade Policy Analysis – A Handbook*. Cambridge University Press.

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

Stanley, T. D., 1998. New Wine in Old Bottles: A Meta-Analysis of Ricardian Equivalence. *Southern Economic Journal*, 64(3): 713-727.

Thorbecke, E., 2001. *The Social Accounting Matrix: Deterministic or Stochastic Concept?* Paper prepared for a conference in Honour of Graham Pyatt's retirement, the Hague, Netherlands, November 29th.

van der Mensbrugghe, D., 2000. *Single Country CGE Models Using GTAP Version 5 and their Implementation in Excel*. Working Paper, Development Prospects Group, the World Bank, Washington D.C.

Weichselbaumer, D. and Winter-Ebmer, R., 2005. A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys*, 19(3): 479-511.

Wood, S., 2006. *Generalized Additive Models- an Introduction* with R. Chapman & Hall, London.