Modelling the Dynamics of Market Shares in a Pooled Data Setting

Nowak-Lehmann D., Felicitas; Herzer, Dierk; Vollmer, Sebastian; Martínez-Zarzoso, Inmaculada

Modelling the Dynamics of Market Shares in a Pooled Data Setting

Journal:	*Applied Economics*
Manuscript ID:	APE-07-0026.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	02-Jun-2008
Complete List of Authors:	Nowak-Lehmann D., Felicitas; Ibero-Amerika Institut, Economics Herzer, Dierk; University, Economics Vollmer, Sebastian; University, Economics Martínez-Zarzoso, Inmaculada; Institute of Applied Economics, Economics
JEL Code:	F14 - Country and Industry Studies of Trade < F1 - Trade < F - International Economics, C13 - Estimation < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, C23 - Models with Panel Data < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods
Keywords:	dynamic panel data model, standard autoregressive distributed lag model, Three-Stage Feasible Generalized Least Squares estimation, market shares

Modelling the Dynamics of Market Shares in a Pooled Data Setting

By

Felicitas Nowak-Lehmann D.: Ibero-America Institute for Economic Research, University of Goettingen, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany, ph.: +49 551 397487, fax: +49 551 398173, fnowak@uni-goettingen.de (corresponding author)

Dierk Herzer: Johann Wolfgang Goethe-University, Frankfurt, Germany DierkHerzer@wiwi.uni-frankfurt.de

Sebastian Vollmer: Ibero-America Institute for Economic Research, University of Goettingen, Germany, Sebastian.Vollmer@wiwi.uni-goettingen.de

Inmaculada Martínez-Zarzoso: Department of Economics, Applied Economics, Universidad Jaume I, Castellón, Spain, martinei@eco.uji.es

Abstract

The objective of this paper is twofold: first, to study the applicability of the widely used Autoregressive Distributed Lag Model (ARDL) in a pooled data setting. Second, it is to analyze Chile's market shares in the EU in the period 1988–2002, pointing to application problems that might jeopardize the model and searching for estimation methods that deal with the problem of inter-temporal and cross-sectional correlation of the disturbances. To estimate the coefficients of the ARDL model, Feasible Generalized Least Squares (FGLS) is utilized within the Three Stage Least Squares (3SLS) framework. A computation of errors is added to highlight the susceptibility of the model to problems related to the underlying model assumptions.

Keywords: dynamic panel data model, standard autoregressive distributed lag model; pooled Three-Stage Feasible Generalized Least Squares estimation (3SFGLS), market shares

JEL: F14, C13, C23

Modelling the Dynamics of Market Shares in a Pooled Data Setting

1. Introduction

In this paper, a standard Autoregressive Distributed Lag Model (ARDL) is utilized to estimate the dynamics of Chile's market shares in the EU market. This dynamic model has been adapted from studies of Balestra and Nerlove (1966), Baltagi and Levin (1986), Arellano and Bond (1991), Blundell et al. (1992), Islam (1995), and Ziliak (1997), Kim et al. (2003) among others. Cable (1997) applied an ARDL to market share behavior and mobility in the UK daily newspaper market. A common feature of all these studies (and many more of this kind) is that the dynamic relationship between dependent and independent variables is captured by a lagged dependent variable, thus leading to an autoregressive distributed lag model. This is "the" standard dynamic model that is applied to panel data, as described in Baltagi (2005).

The main aim of this paper is to examine the applicability of the ARDL from both a theoretical and an empirical point of view. From a theoretical perspective, we analyze the structure and origin of this widely used autoregressive distributed lag model. From an empirical perspective, we illustrate estimation problems of the ARDL with an empirical application to Chile's market shares in the EU market. We differentiate among three types of caveats that can be lodged. The first is theoretical and deals with the underlying assumptions of the ARDL and the underlying geometric lag structure. The second caveat deals with the time series properties of the data and the autocorrelation problem present in most panel data sets. The third caveat centers around the endogenity of the lagged dependent variable on the right-hand side and the endogenity of standard instrumental variables in the presence of serial autocorrelation.

The first type of problem arises because the standard ARDL is derived from a geometric lag model, which presumes that all right-hand side variables impact on the dependent variable in exactly this geometric form (Koyck, 1954). The reason for transforming the geometric lag model into an ARDL is that the former is non-linear in its parameters. Non-linearity in the parameters used to be considered a potential problem for estimation. Today, however, modern computer software makes it possible to apply non-linear least squares to the geometric lag model so that this transformation (Koyck transformation) can be regarded as superfluous. Nonetheless, the simple or standard ARDL continues to be "the" preferred dynamic model since it neatly summarizes the impact of all regressors (lagged and unlagged) in just one variable: the lagged dependent variable. Deriving the ARDL from the geometric lag model, however, reveals how restrictive the autoregressive ARDL can potentially be.

The second type of problem is due largely to the non-stationarity of the data entering the panel analysis. Non-stationarity usually leads to serial correlation, a problem that has to be dealt with if present. Panel unit root tests and panel autocorrelation tests must therefore be applied before running regressions to check for the presence of autocorrelated disturbances.

The third type of problem arises only in the presence of the second. Given autocorrelated error terms, additional estimation problems caused by "derived endogenity" appear. The lack of exogenity of the lagged dependent variable and/or standard instrumental variables is the logical consequence of serial correlation. To tackle these estimation problems, the dynamic pooled data model is estimated by the Three Stage Least Squares (3SLS) method, in combination with Feasible Generalized Least Squares (FGLS) and Seemingly Unrelated Regression (SUR) to deal with the problems of endogenity and autocorrelation of the residuals across countries and over time.

With those inherent problems in mind we apply the ARDL to pooled data (Chile's market shares in different EU countries in the period 1988–2002), using the necessary caution in doing so. In this applied economics part, we also conduct an error analysis. To our knowledge, this is the first study to examine the preconditions, the applicability and estimation problems of the ARDL in a panel or pooled data setting.

The paper is set up as follows. In Section 2, applicability issues and estimation problems of the ARDL are discussed. The derivation of the model and the underlying assumptions of the ARDL are analyzed in Section 2.1, and in Section 2.2, combined estimation techniques are proposed to solve at least some of the estimation problems arising in a pooled data setting. In Section 3, we set up a simple dynamic market share model for Chilean exporters to the EU and study the empirical applicability of the ARDL. Section 4 presents our estimation results and an error analysis. Section 5 concludes.

2. The ARDL model in a panel/pooled data setting

There are different possibilities for modeling dynamic relationships. To save parameters, specific dynamic models[1] have been developed, e.g., the Gamma lag model, which is presumptive in form, as well as the polynomial lag model and the transfer function model, which are quite flexible in that they make it possible to model any reaction pattern that can be shaped by a polynomial (polynomial lag model) or by a ratio of two polynomials (transfer function model or autoregressive distributed lag model)[2]. However, the most widely used dynamic model is the first-order autoregressive distributed lag model with only a lagged dependent variable capturing the impact of current and lagged explanatory variables. For simplicity this will be called a simple autoregressive distributed lag model (ARDL).

2. 1 The simple ARDL as the standard dynamic model

[1] The workings of the models, their advantages and disadvantages can be found in Greene (2000) and Nowak-Lehmann D. (2004).
[2] The transfer function model is the most general formulation of the ARDL in that it allows for impacts of the lagged dependent variable and of all lagged independent variables.

The simple ARDL model (see eq. (1) below) has become the most popular of all the dynamic models since the lagged reaction between dependent and explanatory variables is captured in a single parameter, which is known as the adjustment parameter λ. This parameter expresses the reaction between y_{it} and $y_{i,t-1}$ explicitly and the reaction between y_{it} and $x_{1it},\dots, x_{1i,t-k}, \dots,$ $x_{pit},\dots, x_{pi,t-k}$ implicitly. λ "summarizes" the impact of all p independent variables. In a panel or pooled data context, the ARDL can exist in two forms: the random-effects (RE) form and the fixed-effects (FE) form. The Panel ARDL (see Baltagi, 2005) is of the following form:

$$y_{it} = a + b_0\, x_{1it} + c_0\, x_{2it} + \dots + q_0\, x_{pit} + \lambda\, y_{i,t-1} + u_{it} \qquad (1)$$

with i=1,…., N; t=1,…., T; a being a common intercept, b_0, c_0, …, q_0… being the impact multipliers and λ = adjustment parameter and $u_{it} = \mu_i + \upsilon_{it}$. The two components of the error term are independent of each other and among themselves so that μ_i ~IID(0; σ_μ^2) and υ_{it} ~ IID(0; σ_υ^2) hold. This implies that autocorrelation of υ_{it} is assumed away.

To estimate the panel ARDL, various GMM estimators based on first differencing were developed by Holtz-Eakin, Newey and Rosen (1988), Arellano-Bond (1991), Arellano and Bover (1995) and Blundell-Bond (1998). Keane and Runkle (1992) created a first difference estimator that works differently from the GMM estimators and utilizes the forward-filtering procedure to find adequate instruments for $\Delta y_{i,t-1}$ that are not correlated with the error terms. These methods ensure unbiased and consistent estimates if the υ_{it} are not autocorrelated.

However, the ARDL can be accompanied by two categories of problems that have not been dealt with in the dynamic panel literature. The first is related to the underlying model and its very restrictive assumptions (Section 2.1.1). The second has to do with the time series properties of the variables and the high likelihood of autocorrelation of the disturbances, that renders the application of standard GMM techniques obsolete (Section 2.2).[3].

2.1.1 The underlying model and the ARDL assumptions

In 1954, Koyck showed that the geometric lag model (eq. (2)) below is the underlying model of Eq. (1). For the FE case we obtain:

$$y_{it} = a_i + b_0 \lambda^0 x_{1it} + \ldots + b_0 \lambda^k x_{1it-k} + \ldots\ldots +$$

$$q_0 \lambda^0 x_{pit} + \ldots + q_0 \lambda^k x_{pit-k} + \upsilon_{it} \qquad (2)$$

$0 < \lambda < 1$; and $b_k = b_0 \lambda^k$ denotes the impact of a change that happened k periods ago

λ is the same for all regressors $x_{1it}, \ldots, x_{pit}$

In the geometric lag model all explanatory variables (x_{1it}, $x_{2it}, \ldots, x_{pit}$) have a geometrically declining impact on the dependent variable y_{it}, in such a way that changes in the more distant past have a more minor impact than changes in the more recent past (see Illustration 1). In a multivariate dynamic regression model, all explanatory variables (x_{1it}, $x_{2it}, \ldots, x_{pit}$) have to impact on y_{it} in exactly the same geometric way, with the same λ. This leads to a further complication in the ARDL in the multivariate regression model compared to the bivariate regression model, where only one regressor (x_{it}) would have to have this geometric impact on y_{it}.

[Illustration 1 about here]

However, there are many instances in which the assumption of a geometric lag itself will not be fulfilled. This will be true especially when reaction lags are present and when changes in the current and the preceding periods therefore have a lower impact than changes in earlier periods. In such cases, a better option is a polynomial lag model, which allows us to estimate any lag structure that can be depicted by a polynomial of order 1, 2,…, p.

[3] Standard GMM utilizes lagged variables as instruments. This leads to biased estimates in the presence of

2.1.2 Deriving the ARDL by means of the Koyck transformation

Koyck (1954) was the first to show how a geometric lag model can be transformed into an ARDL. This transformation is called Koyck (lag) transformation. Eq. (2) lagged by one period and multiplied through with λ gives,

$$\lambda y_{it-1} = \lambda a_i + b_0 \lambda^0 \lambda x_{1it-1} + \ldots + b_0 \lambda^k \lambda x_{1it-k-1} + \ldots\ldots +$$

$$q_0 \lambda^0 \lambda x_{pit-1} + \ldots + q_0 \lambda^k \lambda x_{pit-k-1} + \lambda \upsilon_{it-1} \quad (3)$$

Eventually we obtain the ARDL (Koyck lag formulation of geo lag model, eq. (4)) by substracting (3) from (2) and by ignoring the terms $b_0 \lambda^{k+1} x_{1it-k-1}, \ldots, q_0 \lambda^{k+1} x_{pit-k-1}$.

$$y_{it} = a_i^* + b_0 x_{1it} + c_0 x_{2it} + \ldots + q_0 x_{pit} + \lambda y_{it-1} + v_{it} \quad (4)$$

with $a_i^* = a_i(1-\lambda)$ and $a_i = a_i^*/(1-\lambda)$ and $v_{it} = \upsilon_{it} - \lambda \upsilon_{it-1}$; Note that a_i is from eq. (2).

The advantage of the ARDL is that the number of lag coefficients to be estimated reduces to $b_0, \ldots, q_0$ and λ and that all impact-coefficients b_1, q_k can be easily computed according to the general formula: $b_{lag} = b_0 \lambda^{lag}$, thus generating: $b_1 = b_0 \lambda^1$, $b2 = b_0 \lambda^2$. A further advantage of the ARDL is that it is linear in its coefficients, thus allowing the application of linear estimation techniques.

A severe shortcoming of the ARDL is that the estimators will be inefficient and biased (even inconsistent) in the presence of autocorrelation of the disturbances (Kelejian and Oates, 1981). However, also the problem of neglecting $b_0 \lambda^{k+1} x_{1it-k-1}$ can lead to huge errors (estimation mistakes) if λ is relatively large and the maximum lag, k, is short. A short lag length might be a problem when working with annual data and not so much when working with daily or weekly data. This point will be elaborated and illustrated in the empirical section (Section 4, Table 3).

---

autocorrelation of the disturbances.

2.1.3 Popularity of the ARDL in practice

The question, then, is why the ARDL is so popular if it is so restrictive and burdened with estimation problems. There are two possible answers. First, the Koyck lag transformation is a very simple formulation, thus, the applied economist does not have to worry about the lag length, at least not initially.[4] It is also the "standard" dynamic model, leading the researcher to believe it is very flexible and general in the pattern described. The second reason why model (4) tends to be used rather than model (2) is historical. The popularity of the ARDL seems to be due to the econometric software limitations that existed several decades ago, where non-linear least squares estimation was avoided in favor of linear least squares estimation, the feasible method with the software of the time. Moreover, the parameter-saving estimation technique of the ARDL is an attractive feature.

Compared to the ARDL, the original model (2) would be a slightly better option (than eq. (4)) for the applied economist since it could at least allow for different geometric lags for different regressors thus leading to,

$$y_{it} = a_i^* + b_0 \lambda^0 x_{1it} + \ldots + b_0 \lambda^k x_{1it-k} + \ldots\ldots +$$

$$q_0 \delta^0 x_{pit} + \ldots + q_0 \delta^k x_{pit-k} + v_{it} \qquad (5)$$

Eq. (5) could be estimated by Non-linear Least Squares (NLS).

In case of autocorrelation of the disturbances, Feasible Generalized Least Squares (FGLS) can be applied to eq. (5) to correct for autocorrelation. With no lagged dependent variable on the right-hand side, "derived" endogenity of the right-hand variables would not occur. The problem of "derived endogenity" will be explained in Section 2.2.

[4] Only in the error analysis will the lag length k be of relevance.

2.2 Estimation problems arising in an ARDL with pooled data

In the dynamic panel data literature, emphasis has been placed on first-differencing the series to eliminate the individual effects and in finding instruments that are not correlated with the error terms in first differences. Autocorrelation of the disturbances has been assumed away. This focus certainly has to do with applying the ARDL to **panel** data.[5] Panel data mostly correspond to data with large numbers of cross-sections, with variables held in single series in stacked form. The dynamic panel data models usually use a small number of observations over time (e.g. T=2 or T=3), so that in the Generalized Method of Moments (GMM) estimation, the number of lagged instruments and moment conditions (T*(T-1)*(p/2) does not get too large, with p being the number of regressors.

However, establishing a dynamic structure with only a small number of observations over time is not a satisfying approach since building a dynamic model based on a very small number of observations over time can only capture the true dynamics "by chance". If possible, one should work with a longer time span (pooled data setting). In this setting, the number of observations over time is usually larger than the number of cross-sections (T>I) and therefore the time series properties of the variables become relevant. All the time series problems must be dealt with and the time series properties of the variables must be scrutinized very carefully in order to avoid running spurious regressions. This problem can occur when the series in a regression model follows a deterministic trend or a difference-stationary (I(1), I(2),…) process. It was discussed extensively by Granger and Newbold (1974) and led to the development of stationarity tests (unit-root tests) such as the Augmented Dickey Fuller test (1979), the Phillips-Perron test (1998), the Kwiatkowski, Phillips, Schmidt, and Shin test (KPSS, 1992), the GLS-detrended Dickey-Fuller test (Elliott, Rothenberg, and Stock, 1996),

[5] The ARDL and its estimation techniques are discussed in Chapter 8 of Baltagi (2005), "Dynamic Panel Data Models".

the Elliott, Rothenberg, and Stock Point Optimal test (ERS, 1996), and the Ng and Perron test (NP, 2001).

Scrutinization of the series implies checking whether the series of the regression model have a memory, i.e. whether y_{it}, x_{1it},…, x_{pit} are determined by their past values. As is well known from the time series literature, the unit root tests check whether a series (see eq. (6)-(8)) is non-stationary, whether it has a unit root, with $|\rho_i| \geq 1$ in the H_0-hypothesis (Davidson and MacKinnon, 1993, Hamilton, 1994, and Hayashi, 2000).[6]

$$y_{it} = \rho_{yi}\, y_{it-1} + u_{it} \tag{6}$$

$$x_{1it} = \rho_{x1i}\, x_{1it-1} + u_{it} \tag{7}$$

$$x_{pit} = \rho_{xpi}\, x_{pit-1} + u_{it} \tag{8}$$

In the last ten years, enormous progress has been made in the field of panel/pool unit root tests. There are two types of panel/pool unit root tests. One type assumes panel homogeneity (common unit root processes for all cross-sections). These tests were developed by Levin, Lin and Chu (2002), Breitung (2000), and Hadri (1999). The second type of tests allow for panel heterogeneity. The Im, Pesaran, Shin test (IPS test, 2003), the Fisher-type tests using ADF and Phillips-Perron test (Maddala and Wu, 1999; Hadri (2000), Choi (2001), Hadri and Larsson (2005)) are based on individual unit roots or coefficients ρ_i for each cross-section. Unit root tests that allow for panel heterogeneity should be given priority over common unit root tests for obvious reasons. We will apply the IPS test in the empirical part.

If the series in the regression model has a strong memory of the past, there is a very high likelihood that the omitted variables, which are lumped together in the error term v_{it}, will have a strong memory of the past as well. They need not possess a unit root (be non-stationary with a ρ close to one), but the probability that the error terms will be autocorrelated is high, i.e.

$v_{it} = \rho_i\ v_{i,t-1}$, with ρ_i being significantly different from zero.

In a panel or pooled data setting, AR terms (AR(1), AR(2) etc.) are plugged into the regression to test for autocorrelation of the disturbances. Alternatively, one can run a pooled regression of the ARDL, compute the residuals, regress the residuals on the residuals of the previous period(s)[7] and then test whether the autocorrelation coefficient is significant. If autocorrelation is detected, feasible generalized least squares (FGLS) is applied.

However, autocorrelation leads to another estimation problem not solved by GMM.

According to eq. (4) we will always obtain unavoidable correlation between v_{it} and y_{it-1} , if the disturbances are autocorrelated. Due to autocorrelation of the residuals, y_{it-1} will become endogenous and must be instrumented. Typical GMM procedures such as those summarized in Baltagi (2005) cannot be applied since lagged y_{its} (both in levels or in differences) will also be correlated with the error term through $\upsilon_{i,t-1}$.

If there is endogeneity, we suggest running Three-Stage Least Squares (3SLS)[8]. This will be shown in Section 4.

3. The ARDL model with pooled market share data

Despite the ARDL model's restrictiveness, we apply it to Chilean pooled data with a critical attitude and an acute sense of caution, checking whether the model's assumptions are fulfilled and applying combined estimation techniques that allow us to control for autocorrelation, heteroscedasticity and cross-sectional correlation of the residuals.

From an applied economist's point of view, the objective of this paper is to analyze Chile's market share in the EU market on a sectoral level in the period 1988–2002, also applying the necessary panel/pooled time series techniques. The ARDL model is built on six cross-sections

[6] For simplicity, no constant and no trend are included in the equations.
[7] One must test for first-order, second-order etc. autocorrelation of the residuals. When working with annual data, first-order correlation is most common.
[8] or a modified GMM with instruments different from the ones suggested by GMM.

(EU countries) and fifteen annual observations for Chile's seven most important export sectors (fish, fruit, wine, ores, wood, wood pulp, and copper).

Market shares in a specific sector (s) are computed as a ratio of Chile's sectoral exports (X in the numerator) and EU country i's imports from the world $M_{.i} = M_{EUi}+M_{non\text{-}EUi}$ (in the denominator). Due to unsubstantial trade volumes, we consider only Chile's market shares in France (FRA), the Netherlands (NDL), Germany (DEU), Italy (ITA), UK (GBR), and Spain (ESP). Market shares are computed for seven sectors at the two-digit HS levels, namely fish (03), fruit (08), beverages (22), ores (26), wood (44), wood pulp (47) and copper (74). Sources of the data and generation of the data are described in Appendix 1 and the development of market shares is depicted in Figures 1 to 7 in Appendix 2. The selection of sectors and competitors is based on COMEXT and TradeCAN data. The period covered goes from 1988 to 2002. Thus, we obtain a maximum of six cross-sections and 15 years, resulting in a maximum of 90 observations per sector. The number of observations varies depending on the sector studied.

3.1 The market share ARDL model

There are two approaches to modeling market shares: according to the first, market shares are basically purely stochastic, and according to the second, they are influenced by hard economic factors such as prices, marketing expenditures, the number and strength of competitors, etc. To model market shares, Sutton (2004) chooses an eclectic approach. Favoring the idea of building a stochastic model, he expands the model to include industry-specific features (e.g., a strategic representation of firms' competitive responses to market share changes). However, it has to be kept in mind that strategic behavior is very often intrinsically unobservable.

Cable (1997) proposes modelling market shares according to the second approach using an autoregressive distributive lag model (ARDL)[9]. He selects a first-order autoregressive model with a one-period lagged endogenous variable[10], in which prices and advertising shares are the explanatory variables for the UK's national daily newspapers.

We modify Cable's model as follows: according to our model (see eq. (9)), market shares are determined by Chile's and its main competitors' relative prices in the EU countries and an unobserved variable such as strategic behavior[11]. Price competitiveness[12] is considered a decisive determinant of Chile's market shares, since Chile's most successful exports are rather homogeneous products (fish, fruit, beverages, ores, copper, and wood and products thereof). Accordingly, Chile's market share in a specific sector is determined by Chile's price advantage (in terms of EU-Chilean producer prices and EU protection) and Chile's competitors' price advantage on the EU market.

$$lshw_{it} = \alpha_i + \beta_0\, lreer_{it} + \gamma_0\, lreer^*_{it} + \lambda\, lshw_{it-1} + v_{it} \qquad (9)$$

$$lshw_{it} = \alpha_i + \beta_0\, lreer_{it} + \beta_0 \lambda^{k+1} lreer_{ist-k-1} + \gamma_0\, lreer^*_{it} + \gamma_0 \lambda^{k+1} lreer^*_{ist-k-1}$$

$$+ \lambda\, lshw_{it-1} + v_{it} \qquad (10)$$

Eq. (9) is the simple ARDL and eq. (10) is the complete Koyck transformation without dropping the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer^*_{ist-k-1}$,

where

i = 1, 2,…, 6 represents the cross-sections: FRA, NDL, DEU, ITA, GBR and ESP (according to World Bank abbreviations);

[9] First-order autoregressive model.
[10] There are two types of autoregressive distributed lag models: the geometric lag model and the transfer function model, also known as ARMAX model (for a good description, see Greene, 2000)
[11] Since strategic behavior is difficult to model, we assume that strategic behavior and sector-specific characteristics are incorporated into the residuals of the regression model
[12] We believe that exchange rates, cost differentials, tariffs and subsidies are important 'hard' factors explaining market shares over time..

t = 1988, 1989, …, 2002 are years (annual observations)

s = 03, 08, 22, 26, 44, 47 and 74 are the sectors (according to the two-digit HS classification)

$lshw_{ist}$ stands for Chile's market share in EU country i in sector s at point t. $lreer_{ist}$ is Chile's real effective exchange rate, prevailing in EU country i and in sector s and $lreer*_{ist}$ is Chile's competitor's (*) real effective exchange rate, prevailing in country i and in sector s. Since *lreer* and *lreer** are in price quotation, we expect *lreer* to have a positive impact and *lreer** to have a negative impact.

We estimate eq. (9) as a fixed effect model allowing for cross-section-specific intercepts (α_i). This model could be applied in its unrestricted form by estimating cross-section-specific slope parameters for $lreer_{it}$, $lreer_{it}*$ and $lshw_{it-1}$ (β_{0i}, γ_{0i} and λ_i) but given our limited number of observations in each cross-section, we stick to common slope parameters in all countries. We capture country-specific effects only through cross-section-specific intercepts (α_i) and try to save degrees of freedom by modeling common slope parameters (β_0, γ_0 and λ).

As shown in Section 2.1.1, eq. (9) and eq. (10) are derived from the geometric lag model (eq. (11)).

$$lshw_{ist} = \alpha_{is} + \beta_0 \lambda^0 lreer_{ist} + \ldots + \beta_0 \lambda^k lreer_{ist-k} + \gamma_0 \lambda^0 lreer*_{ist} + \ldots + \gamma_0 \lambda^k lreer*_{ist-k} + v_{ist} \quad (11)$$

As to the coefficients, it is assumed that: $0 < \lambda < 1$ and that λ is the same for all regressors.

It is furthermore assumed that $\beta_{lag} = \beta_0 \lambda^{lag}$, $\gamma_{lag} = \gamma_0 \lambda^{lag}$ and $v_{ist} \sim N(0; \sigma_v^2)$.

Note that eq. (11) assumes not only a geometric reaction of the market share (*lshw*) with respect to relative prices (β_i and γ_i must follow a geometric lag) in all six importing countries i under investigation, but it assumes exactly the same (as measured by λ_i) geometric reaction of *lshw* with respect to changes of all the regressors (both *lreer* and *lreer**). In our case, as well as in many other studies using the ARDL, the above assumption cannot be justified by the data for all regressors. Also, the specific geometric reaction does not always apply to all countries under study. These issues would become even more crucial with an increasing number of cross-sections and with some more explanatory variables in the model (a model with, e.g., 100 countries and five regressors).

Therefore, before applying our data to the ARDL, we examined the cross-correlations between the dependent and the independent variables[13] (12 per sector, 84 cross-correlations in total). With the help of cross-correlations, the dynamics of the model (the lag structure between dependent and independent variable) can be studied. The cross-correlations indicate that the geometric lag assumption is not fulfilled in the majority of cases and that the maximum lag length is between two and three years.

3.2 Estimation techniques in the presence of non-stationary market share data

Assuming for the moment that the underlying assumptions with respect to the geometric lag of the ARDL model are fulfilled, the time series properties of the data are checked and a test of autocorrelation of the disturbances is applied. We will first have a look at market shares and test their persistence by means of panel/pooled unit root tests[14]. If market shares turn out to be stationary (I(0))[15], this will indicate that they are robust and persistent during the period 1988–2002. However, if they result to be non-stationary[16], then we will conclude that the Schumpeter 'hypothesis' cannot be rejected by the 1988–2002 data. Of course, our time period is too short to draw conclusions on whether the Schumpeter hypothesis is valid in the long run.

[13] These cross-correlations show the reaction pattern between the dependent and the independent variables very clearly and should precede the building of any dynamic models. The 84 cross-correlations are available from the authors upon request.

[14] According to Schumpeter market shares or leadership positions are transient. Schumpeter labels those leadership positions that arise from invention and innovation 'temporary monopolies'. According to Alfred Chandler market shares are robust over time and leadership tends to persist for a 'long' time. (2002 Japan Conference, 2005).

[15] Chandler's view.

[16] Schumpeter's view.

3.2.1 Testing the time series properties of the market share data

We proceed in several steps. In the first, we test the time series properties of the data (all in natural logs). All series, i.e., market shares (*lshw*), Chile's real effective exchange rate (*lreer*) and Chile's competitors' real effective exchange rates (*lreer**) for all country pairs are subjected to tests of non-stationarity (panel unit root tests). This procedure is applied to all seven sectors under investigation neglecting the possible existence of structural breaks in the series because neither fundamental, abrupt changes in economic policy, nor any major exogenous shocks were detected in the period 1988–2002.[17]

In the statistical analysis we allow for different unit root processes in the panel, i.e. cross-section-specific (country-specific) unit roots. We apply the Im, Pesaran, and Shin (2003) panel unit root test to all series considering the possibility of individual unit roots of our panel data. According to Table 1, all variables (*lshw*, *lreer*, and *lreer**) are non-stationary, integrated of order one (I(1)) with a p-value of 0.00 (exception: *lrpcopper* with p = 0.02). The critical t-bar value for $\alpha=1\%$ is -2.48. Of course, we have to be cautious in interpreting the results, since unit root tests generally tend to falsely accept the unit root null in small samples. On the other hand, we can already conclude from the plots of the data (Appendix 2) that the market shares exhibit non-stationary behavior.

[Table 1 about here]

With respect to market shares, this finding supports Schumpeter's view that gains in market shares are temporary. Monopolistic positions have to be defended; otherwise they are lost quickly. This view seems to apply especially to the fish, fruit, beverages, ores, and the copper sector. Market shares appeared more stable in the wood sectors (44 and 47) (see Figures 5-6 in Appendix 2), but are non-stationary according to the tests.

[17] The economic policy of the Pinochet government was continued under the governments of Aylwin, Frei and Lagos. Consequently, the time series display no sign of a significant structural shift.

3.2.2 The FGLS approach versus panel cointegration and error correction approaches

Since all variables are I(1), one could proceed with cointegration analysis and panel cointegration tests (Pedroni, 1999; Pedroni, 2004). However, cointegration is a long-term concept that is not applicable to our short time span. Moreover, with fifteen annual observations, the power of panel cointegration tests would be too low. But cointegration analysis is not the only approach that deals with non-stationary series and yields unbiased and efficient estimates in a dynamic model. FGLS is another possibility, as is known from time series analysis. Therefore, we exploit the special suitability of FGLS for estimating dynamic models with panel data (see Stock and Watson, 2003).

In a panel/pooled analysis setting, FGLS works in analogy to the time series setting. The idea remains the same: non-stationarity of the series in a regression equation is reflected in the autocorrelation ρ of the residuals over time. Annual data usually shows first-order autocorrelation, and this is the case in our sample as well.[18]

The procedure will be described below by abstracting from sectors for a moment. We estimate ρ_{ik} of eq. (12) below, after having computed the residuals $\hat{v}_{it}$ from the ARDL model (eq. (9))

$$\hat{v}_{it} = \Sigma_{k=1}^{K} \rho_{ik} \hat{v}_{it-k} + e_{it} \tag{12},$$

with $e_{it} \sim N(0; \sigma_{ei}^2)$ and k = 1, 2,…K number of lags. Autocorrelation of the residuals is strongly related to the presence of non-stationary series. The autocorrelation coefficient ρ_{ik} [19] in a way captures the autoregressive processes (expressed by ρ_{ik}', ρ_{ik}'' and ρ_{ik}''') prevailing in the series (see equations (13)-(16)).

In theory we have:

---

[18] ρ is usually well below 1 so that first differencing is a very rough method to get rid of stationarity.

$$lshw_{it} = \Sigma_{k=1}^{K} \rho'_{ik}\, lshw_{it-k} + e'_{it} \quad (13)$$

$$lreer_{it} = \Sigma_{k=1}^{K} \rho''_{ik}\, lreer_{it-k} + e''_{it} \quad (14)$$

$$lreer^*_{it} = \Sigma_{k=1}^{K} \rho'''_{ik}\, lreer^*_{it-k} + e'''_{it} \quad (15)$$

$$lshw_{it-1} = \Sigma_{k=1}^{K} \rho^{iv}{}_{ik}\, lshw_{it-k-1} + e^{iv}{}_{it-1} \quad (16)$$

Note that FGLS uses a common $\hat{\rho}_k$ (with k signaling the order of autocorrelation) in equations (13)-(16) and transforms the variables correspondingly.

The FGLS method is applied in three steps: first, eq. (9) is estimated by SUR and the residuals are computed. Second, the order (first-order, second-order, or p-order) of autocorrelation $\hat{\rho}_k$ is estimated applying SUR and significance is tested in eq. (12). First-order autocorrelation of the type $\hat{v}_{it} = \hat{\rho}_{i1}\, \hat{v}_{it-1}$ turns out to be present and dominant. $\hat{\rho}_{i1}$ expresses first-order autocorrelation, henceforth to be called $\hat{\rho}$. Third, if only first-order autocorrelation is present (as in our case), the variables of eq.(9) are transformed into

$$lshwz_{it} = lshw_{it} - \hat{\rho}\; lshw_{it-1} \quad (17)$$

$$lreerz_{it} = lreer_{it} - \hat{\rho}\, lreer_{it-1} \quad (18)$$

$$lreerz_{it}^* = lreer_{it}^* - \hat{\rho}\, lreer_{it-1}^* \quad (19)$$

$$lshwz_{it-1} = lshw_{it-1} - \hat{\rho}\, lshw_{it-2} \quad (20)$$

and $\varepsilon_{it} = \hat{v}_{it} - \hat{\rho}\; \hat{v}_{it-1}$ (21)

thus generating variables in soft or quasi-first differences. Eq. (9) is then estimated on the basis of the transformed variables, leading to eq. (22) (see Stock and Watson, 2003).

$$lshwz_{it} = \alpha_{is}(1-\hat{\rho}) + \beta_0\; lreerz_{it} + \gamma_0\; lreerz_{it}^* + \lambda\; lshwz_{it-1} + \varepsilon_{it} \quad (22)$$

[19] Which is to be estimated since it is unknown.

3.2.3 Autocorrelation of the disturbances as a possible result of non-stationarity

Non-stationarity of the series (see Table 1) is usually linked to first-order correlation of the residuals (see Table 2). An AR term in the equations can indicate this problem in a pooled data setting where the Breusch-Godfrey LM test is not feasible.

Even though serial correlation in dynamic panel data models is only rarely dealt with in the econometric literature, the studies by Hujer, et. al. (2005), Kim et al. (2003), Sevestre and Trognon (1996) and Keane and Runkle (1992) dwell on this issue. Keane and Runkle (1992) and Kim et al. (2003) use the forward-filtering 2SLS method (KR estimate). This method pretends serial correlation to be equal to one, which is a very rough estimate. Kim et al. (2003) refine the KR method. We, in contrast, estimate the extent of serial correlation in the sample (our $\hat{\rho}_{ik}$)[20] and then transform the variables correspondingly (in soft or quasi-first differences).

Hujer et al. (2005) assume that the residual term follows a moving average process (eg. MA(1), MA(2)). According to our data, however, the residual terms clearly follow an AR(1) process and not an MA(1) process. Panel/pooled analyses with macroeconomic data usually show unit roots in the series and are therefore characterized by an autoregressive error process. For this reason, time series tests on the series and the residuals are a must before starting estimation of the model.

Moreover, as we have seen before, the advantage of having a linear model comes at the cost of having a lagged endogenous variable that is correlated with the disturbance term due to autocorrelation. When a lagged endogenous variable appears on the right-hand side of a regression equation (as in eq. (9)) and when the disturbances are autocorrelated, the lagged endogenous variable will automatically be correlated with the disturbance term and thus become endogenous. The endogenity problem of the lagged dependent variable ($lshw_{it-1}$), which is caused by first-order AR correlation of the residuals due to non-stationarity of the series., can be effectively tackled by the Three-Stage Least Squares[21] technique utilizing 3SFGLS. Modern computer programs allow one to generate the variables in soft first differences directly by adding, e.g., an AR(1) term for first-order autocorrelation and to simultaneously apply methods that control for the endogeneity of the regressors.

4. Estimation results for the market share ARDL

For each sector, separate panel ARDLs (applying eq. 9) have been run over the time period 1988–2002, with the EU countries acting as cross-sections in the panel analysis. To control for autocorrelation, FGLS has been combined with a 3SLS routine. The regression results are presented in Table 2. Presence of autocorrelation led us to replace the lagged market share variable and relative prices with variables that are not correlated with the error term and highly correlated with the variables on the right-hand side.

4.1 Regression results (3SFGLS Approach)

The choice of instruments is crucial in order to obtain consistent estimates in any model, also in the market share model. We used an indicator of production capacity in real terms as an instrument for lagged market share ($lshw_{it-1}$), the difference in PPP-income between Chile and the importing country as an instrument for $lreer_{it}$, and the competitor's real exchange rate in a transformation that is generally used in polynomial lag models as an instrument for $lreer^*_{it}$. Table 2 summarizes the impact of price competitiveness on market shares estimated by Three Stage Feasible Generalized Least Squares (3SFGLS).

---

[20] In FGLS, the unknown serial correlation coefficient is estimated as described in Section 2.

[21] Three-Stage Least Squares (3SLS) technique is the SUR version of Two-Stage Least Squares (see EViews 5: User's Guide, 2004, p. 700)

[Table 2 about here]

Under the assumption that the data follow an ARDL model, we find a significant positive impact of increased Chilean price competition on market shares in the fish (03), fruit (08) and ores (26) sector but no significant negative impact of *foreign* price competition on market shares in the seven sectors under study. As to beverages, we find a negative impact of competitive (low) Chilean prices and a positive impact of low foreign prices on market shares.

In the beverages sector (dominated by wine), the quality aspect is supposed to be dominant. FAO statistics (FAO Production Yearbook, 2003; FAO Trade Yearbook, 2003) show that Chile increased its production in the period 1978–2002. Such a production increase, which is usually achieved by intensified irrigation and fertilization, leads to inferior wines at lower prices. The role of prices in the wood (44) and wood pulp (47) sectors might be severely impeded by illegal logging and illegal imports of wood products. Illegal logging distorted official trade flows not only of all timber products (roundwood, sawn wood, veneer, plywood, boards, semi-finished and finished products, and furniture), but also of pulp, paper, printed products and cellulose[22].

Adjustment to the long-run equilibrium was significant in the beverages (22), ores (26), wood (44), wood pulp (47) and copper (74) sectors, whereas no significant adjustment took place in the fish (03) and fruit (08) sectors. However, the results must still be taken with caution, as the error analysis below (Table 3) will show.

4.2 Impreciseness of the incomplete Koyck lag transformation (error analysis of the ARDL)

[22] Illegal logging is estimated to comprise up to 50% of all logging activity in the key countries of Eastern Europe and Russia, up to 94% in the key Asian countries, up to 80% in the key African countries and up to 80% in the key Latin American countries (WWF, 2005; FERN, 2004).

When transforming eq. (11) into the simple ARDL (eq. (9)), which must be considered an incomplete Koyck transformation, an error occurs by suppressing the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer*_{ist-k-1}$. Note that the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer*_{ist-k-1}$ have been dropped to keep the model simple and linear in its coefficients. The shorter the actual lag (k_{max}) and the closer λ (the adjustment parameter) is to one, the larger this error is. The extent of the error becomes intuitively clear by treating eq. (10) as the true ARDL and considering it as the complete Koyck transformation.

The actual error analysis is very straightforward. If the maximum actual lag is k, then the error occurring by dropping the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer*_{ist-k-1}$ is λ^{k+1}. This implies that a maximum lag length of one (two) will lead to an error of λ^2 (λ^3). When working with annual data, one or two year (maximum) lags are very common, such that the danger of committing an error is relatively high.

[Table 3 about here]

We can draw several conclusions from Table 2 and the error analysis in Table 3:

(1) The data do not fit the autoregressive lag model in the fish and in the fruit sector. The λs there carry the wrong sign and are insignificant, since the ARDL requires significant positive λs that lie in an interval $]0;1[$.

(2) The data can be explained by an ARDL in the rest of the sectors by and large since the λs lie in an interval $]0;1[$. However, since we work with annual data where the maximum lag length is usually short (k_{max} = 2 is very realistic according to the cross-correlations), large errors will result in the beverages, ores, and copper sectors, where λ is relatively large and the omission of the terms β_0 λ^{k+1} lreer and γ_0 λ^{k+1} lreer* will therefore result in a large error.

For example, in the copper sector, the error is 64% if k_{max} is 1 and 51% if k_{max} is two. That is, 64% or 51% of the impact of copper prices on the market share in copper is neglected.

(3) Note that the errors are even larger than computed when we have reason to assume that the geometric lag structure does not apply in all instances. Computation of errors in this case would require knowledge of the true model.

To summarize:

On the one hand, we have found that the ARDL estimations have very respectable adjusted R^2 measures and Durbin-Watson (DW) statistics around 2.[23]

On the other hand, the standard errors of the regressions are relatively high. Moreover, the error analysis makes clear that the simple dynamic specification in the form of an ARDL suffers from some drawbacks. The autoregressive lag specification does not seem to apply in the fish and the fruit sector. Statements in the beverages, ores, wood, and copper sectors are subject to relatively large errors due to neglecting the term λ^{k+1}, the impact of changes in prices, and protection[24] in the autoregressive transformation.

Moreover, it has to be noted that the estimation results of 3SFGLS do not fulfill our expectations as far as signs and significance are concerned. This certainly has to do with violated model assumptions but also with the simplicity of the model (we do not control for quality or product innovation). Therefore, the empirical results should not be overemphasized, nor should they be utilized for further analysis.

[23] Even though the DW must be adjusted in the presence of a lagged endogenous, the DW statistic is still able to roughly indicate problems of autocorrelation and misspecification.
[24] All our prices contain sector-specific protection whenever relevant.

5. Conclusions

The ARDL specification can be combined with the FGLS and the SUR technique and is therefore able to deal with several estimation problems resulting from autocorrelation, heteroscedasticity and cross-sectional correlation of the disturbances. Applied to a system of equations, this technique transforms the variables in the regression equation by working with soft differences in the variables and by weighting the regressor matrix with a weight matrix that can control for heteroscedasticity of the variance of the residuals and for cross-sectional correlation of the disturbances (SUR method). The endogenity problem is solved with instrumental variables (IV) in a 3SLS routine. Unlagged exogenous variables are utilized to control for the endogenity problem and to obtain unbiased estimates. Therefore, the 3SFGLS technique is able to produce efficient and consistent estimates if ARDL is the true model.

Violation of the geometric lag assumption is to be expected in particular when working with heterogenous panel data and with multivariate regression models, and will result in inconsistent estimators. Applicability of the ARDL must therefore be tested when working with panel or pooled data. Estimations in the framework of panel error correction models and panel DOLS could be highly advisable, even though these models require much longer time spans to allow for meaningful panel unit root and panel cointegration tests.

Our study has demonstrated that the ARDL model must be applied with caution. First, the geometric lag assumption was not supported overall by the cross-correlations between dependent and independent variables. Second, a maximum lag length of two to three years (also visible in the cross-correlations) can result in substantial estimation errors. Third, non-stationarity of the series leads in general to autocorrelation of the residuals. It renders the utilization of lagged instruments in a standard GMM framework obsolete and requires a search for new instruments, which, however may not be applicable in all cases.

Acknowledgements

We would like to thank Stephan Klasen for his helpful comments and suggestions.

References

Ahn, S. and P. Schmidt (1995), Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics* 68: 5-27.

Anderson, T.W. and C. Hsiao (1981), Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association*; 589-606

Anderson, T.W. and C. Hsiao (1982); Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics* 18: 47-82.

Arellano, M. (1989), A Note on the Anderson-Hsiao Estimator for Panel Data, *Economic Letters* 31: 337-341.

Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies* 58: 277-297.

Balestra, P. and M. Nerlove (1966), Pooling Cross-section and Time-series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica* 34: 585-612.

Baltagi, B.H. and D. Levin (1986), Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation, and Advertising Reconsidered, *Review of Economics and Statistics* 68:148-155.

Baltagi, B. H. (2005), *Econometric Analysis of Panel Data*, Third Edition, Chichester […]: John Wiley & Sons, Ltd.

Blundell, R., S. Bond, M. Devereux and F. Schiantarelli (1992), Investment and Tobin's q: Evidence from Company Panel Data, *Journal of Econometrics* 51: 233-257.

Breitung, J. (2000), 'The Local Power of Some Unit Root Tests for Panel Data', in B. Baltagi (ed.), *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Amsterdam: JAI Press: 161-178.

Breitung, J. and H. Pesaran (2005), Unit Roots and Cointegration in Panels, http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0535.pdf (November 4, 2005).

Cable, J. R. (1997), 'Market Share Behavior and Mobility: An Analysis and Time-series Application, *The Review of Economics and Statistics* Vol. LXXIX (1): 136-141.

Choi, I. (2001), Unit Root Tests for Panel Data, *Journal of International Money and Finance* 20: 249-272.

Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Dickey, D. A. and W. A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* 74.427-431.

Durlauf, S., P. Johnson and J. Temple (2004), Growth Econometrics, Social Systems Research Institute, SSRI Working Papers 18. http://www.ssc.wisc.edu/econ/archive/wp2004-18.pdf (February 6, 2006).

Elliott, G., T. Rothenberg and J. H. Stock (1996), Efficient Tests for an Autoregressive Unit Root, *Econometrica* 64: 813-836.

European Commission (2003), 'Intra- and extra-EU trade, Annual data, Combined Nomenclature, Supplement 2', EUROSTAT, CD ROM of COMEXT trade database.

EU Commission (2005),' The EU Relations with Chile', http://europa.eu.int/comm/external_relations/chile/intro/ (November 24, 2005)

EViews 5: User's Guide (2004), Quantitative Micro Software, LLC, Irvine, CA.

FAO Production Yearbook (2003), Food and Agricultural Organization of the United Nations, Rome, Vol. 57.

FAO Trade Yearbook (2003), Food and Agricultural Organization of the United Nations, Rome, Vol. 57.

FERN, Greenpeace, WWF 2004 :

http://www.panda.org/news_facts/newsroom/news.cfm?uNewsID=17214 (July 28, 2005).

Granger, C. and P. Newbold (1974), Spurious Regressions in Econometrics, *Journal of Econometrics* 2: 111-120.

Greene, W. H. (2000), *Econometric Analysis*, London: Prentice Hall International (UK) Limited.

Hadri, K. (1999), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root in Panel Data with Serially Correlated Errors, *The Econometrics Journal* 3: 148-161.

Hadri, K. (2000), Testing for Stationarity in Heterogeneous Panel Data, *Econometric Journal* 3: 148-161.

Hadri, K. and R. Larsson (2005), Testing for Stationarity in Heterogeneous Panel Data where the Time Dimension is Finite, *The Econometrics Journal* 8(1): 55-69.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Hayashi, F. (2000), *Econometrics*, Princeton, NJ: Princeton University Press.

Holtz-Eakin D., W. Newey and H. Rosen (1988), Estimating Vector Autoregressions with Panel Data, *Econometrica* 56: 1371-1395.

Hujer, R., P. Rodrigues and C. Zeiss (2005), 'Serial Correlation in Dynamic Panel Data Models with Weakly Exogenous Regressors and Fixed Effects, Working Paper, March 9, 2005, J.W. Goethe-University, Frankfurt, Germany.

Im, K., M. Pesaran and Y. Shin (2003), Testing for Unit Roots in Heterogeneous Panels, *Journal of Economterics* 115: 53-74.

Islam, N. (1995), Growth Empirics: A Panel Data Approach, *Quarterly Journal of Economics* 110, 1127-1170.

2002 Japan Conference: A Summary of the Papers, NBER Website; http://www.nber.org/2002japanconf/sutton.html (July 29, 2005).

Judson, R. and A. Owen (1999), 'Estimating Dynamic Panel Data Models: A Practical Guide for Macroeconomists,' *Economic Letters* 65: 9-15.

Keane, M. and D. Runkle (1992), 'On the Estimation of Panel Data Models with Serial Correlation when Instruments are not Strictly Exogenous', *Journal of Business and Economic Statistics* 10(1): 1-9.

Kelejian, H.H. and W.E. Oates (1981), *Introduction to Econometrics. Principles and Applications*, New York: Harper & Row Publishers.

Kim, M.K., G.D. Cho and W. W. Koo (2003), Determining Bilateral Trade Patterns Using a Dynamic Gravity Equation, Agribusiness & Applied Economic Report No. 525, Center for Agricultural Policy and Trade Studies. North Dakota State University. http://www.ag.ndsu.nodak.edu/capts/documents/NovemberNewsletter2003-2-4-04.pdf (March 3, 2006).

Kiviet, J.F. (1995), On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models, *Journal of Econometrics* 68: 53-78.

Koyck, L.M. (1954), *Distributed Lags and Investment Analysis*, Amsterdam: North Holland.

Kwiatowski, D., P. Phillips, P. Schmidt and Y. Shin (1992), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics* 54: 159-178.

Levin, A., Lin, C. F., and C. Chu (2002), Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties, *Journal of Econometrics* 108: 1-24.

Maddala, G.S. and S. Wu (1999), A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics* 61: 631-652.

Ng S. and P. Perron (2001), Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power, *Econometrica* 69(6): 1519-1554.

Nickell, S. (1981), Biases in Dynamic Models with Fixed Effects, *Econometrica* 49: 1417-1426.

Nowak-Lehmann D., F. (2004), Different Approaches of Modeling Reaction Lags: How Do Chilean Manufacturing Exports React to Movements of the Real Exchange Rate?, *Applied Economics* 36(14): 1547-1560.

Phillips, P. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, *Biometrika* 75: 335-346.

OECD (1997), *The Uruguay Round Agreement on Agriculture and Processed Agricultural Products*, OECD Publications, Paris.

Pedroni, P. (1999), Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors, *Oxford Bulletin of Economics and Statistics* 61 (4) Suppl.: 653-670.

Pedroni, P. (2004), Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis, *Econometric Theory* 20: 597-625.

Schmidt, P., S. C. Ahn and D. Wyhowski (1992), Comment, *Journal of Business and Economic Statistics* 10: 10-14.

Sevestre, P. and A. Trognon (1996), 'Dynamic Linear Models', in *The Econometrics of Panel Data. A Handbook of the Theory with Applications,* edited by L. Mátyás and P. Sevestre, pp. 120-144, Dordrecht: Kluwer Academic Publishers, 2nd ed.

Stock, J. (1994), Unit Roots, Structural Breaks, and Trends, Chap. 46 in *Handbook of Econometrics*, Vol. IV, edited by R. Engle and D. McFadden, Amsterdam: Elsevier.

Stock, J. H. and M. W. Watson (2003), *Introduction to Econometrics*, Boston […]: Addison Wesley.

Sutton, J. (2004), Market Share Dynamics and the 'Persistence of Leadership' Debate. The Economics of Industry Group / Suntory and Toyota International Centers for Economics and Related Disciplines, London School of Economics, 37.

World Bank (2002), TradeCAN (Competitiveness Analysis of Nations) 2002 CD-ROM, Washington, D.C.

World Bank (2005), *World Development Indicators,* Data on CD ROM, Washington, D.C.

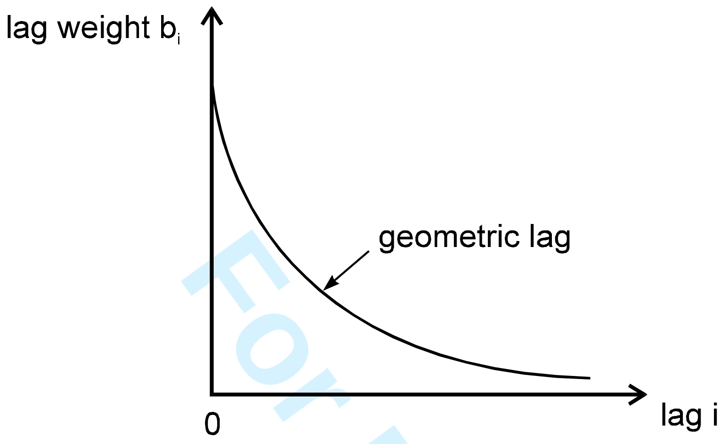
WTO Trade Policy Review, European Union (1995, 1997, 2000), World Trade Organisation, Geneva.

WWF (World Wildlife Fund): EU Imports of Wood Based Products 2002 (2005): http://www.panda.org/about_wwf/wherewework/europe/problems/illegal_logging/ (April 29, 2005)

Ziliak, J. (1997), Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-condition Estimators, *Journal of Business and Economic Statistics* 15: 419-431.

Illustrations and Tables

Illustration 1: Restrictiveness of the underlying assumptions



Note: This reaction pattern must apply to all regressors (x1, x2, x3,…., xp) and at all levels (cross-sections i)! λ is assumed to be the same for all regressors.

Table 1: Results from the Im, Pesaran, Shin (2003) Panel Unit Root Test stating t-bar values

IPS Panel Unit Root Test Based on Individual Unit Roots H_0: Series has a unit root (series is non-stationary)[f]			
Sector 03	**Fish and crustaceans, molluscs**		
	Lshw03	Lreer03	Lreer03*=Lreer03nor
Series in levels Δ Series	-1.81 -4.36	-1.58 -3.42	-1.94 -3.47
Sector 08	**Edible Fruit and nuts**		
	Lshw08	Lreer08	Lreer08*=Lreer08aus
Series in levels Δ Series	-1.68 -5.90	-1.58 -3.42	-2.53 -4.11
Sector 22	**Beverages, spirits and vinegar**		
	Lshw22	Lreer22	Lreer22*=Lreer08saf
Series in levels Δ Series	-1.62 -4.25	-1.58 -3.42	-0.92 -3.34
Sector 26	**Ores, slag and ash**		
	Lshw26	Lreer26	Lreer26*=Lreer26bra
Series in levels Δ Series	-1.29 -4.18	-1.58 -3.42	-2.26 -7.43
Sector 44	**Wood and articles of wood**		
	Lshw44	Lreer44	Lreer44*=Lreer44nor
Series in levels Δ Series	-1.83 -2.80	-1.58 -3.42	-1.94 -3.47
Sector 47	**Pulp of wood**		
	Lshw47	Lreer47	Lreer47*=Lreer47nor
Series in levels Δ Series	-1.68 -2.93	-1.58 -3.42	-1.94 -3.47
Sector 74	**Copper and articles of copper**		
	Lshw74	Lrpcopper[25]	
Series in levels Δ Series	-1.34 -4.22	-1.58 -3.42	----------

Note: lshw = market share, lreer = Chile's real effective exchange rate, lreer* = Chile's competitor real effective exchange rate in sectors 03, 08, 22, 26, 4, 47,and 74.

[f] A trend and an intercept are included in the test equation whenever suggested by the series' graphs.
[25] Lrpcopper serves as an indicator of Chile's real copper production costs. It is used instead of lreer in the market share analysis.

Table 2: Results for the ARDL market share model estimated by 3 SFGLS

	Regression coefficients♣ Equation (2)				Goodness of fit measures♦		
Sector-results	**Impact of lreer** β_{03SLS}	**Impact of lreer*** γ_{03SLS}	**Adjustm. Coeff.** λ_{3SLS}	**AR-term**	**(weighted) R^2adjusted**	**S.E. of regression**	**Durbin Watson stat.**
03 short run	0.82** (0.02)	-0.72 (0.19)	-0.19 (0.20)	0.68*** (0.00)	0.97	1.02	2.15
08 short run	1.82** (0.02)	-0.14 (0.85)	-0.07 (0.70)	0.69*** (0.00)	0.99	1.05	1.99
22 short run	-2.09*** (0.01)	2.01*** (0.01)	0.62*** (0.00)	-0.08 (0.64)	0.98	1.05	2.04
22 long run	-5.50***	5.29***	--------	--------	0.98	1.05	2.04
26 short run	1.83*** (0.00)	0.06 (0.42)	0.70*** (0.00)	-0.29* (0.07)	0.96	1.02	2.06
26 long run	6.10***	0.20	---------	--------	0.96	1.02	2.06
44 short run	0.35 (0.76)	-2.35 (0.13)	0.46*** (0.00)	0.60*** (0.00)	0.94	1.06	2.36
44 long run	0.65	-4.37	----------	---------	0.94	1.06	2.36
47 short run	-1.20*** (0.00)	-0.27 (0.42)	0.37*** (0.00)	0.01 (0.91)	0.99	1.07	1.87
47 long run	-1.90***	-0.43	---------	--------	0.99	1.07	1.87
74 short run	-0.45*** (0.00)	--------	0.80*** (0.00)	-0.07 (0.66)	0.99	1.04	2.16
74 long run	-2.25***	--------	---------	---------	0.99	1.04	2.16

♣ p-vales in brackets.

♦ In 3SLS the adjusted R^2 is negative at times. It is unclear how the goodness of fit measures of the different cross-sections are to be weighted in order to derive an overall goodness of fit measure. Therefore, the figures listed should only signal the trend.

Table 3: Error analysis in the 3SFGLS framework

Sector	Computed adjustment coefficient λ_{3SLS}	Error if k_{max} =1: λ_{3SLS}^{2}	Error if k_{max} =2: λ_{3SLS}^{3}
Fish (03)	-0.19	---	---
Fruit (08)	-0.07	---	---
Beverages (22)	0.62***	0.38	0.24
Ores (26)	0.70***	0.49	0.34
Wood (44)	0.46***	0.21	0.10
Wood pulp (47)	0.37***	0.14	0.05
Copper (74)	0.80***	0.64	0.51

Appendix 1

Description of Data

In the following, the variables: sheu, shnoneu, shw, lreer, and lreer* will be described in original form (not in logs). All data run from 1988 to 2002. Export data (to compute market shares) were taken from EUROSTAT: Intra- and extra–EU trade, Supplement 2, 2003.

In our case, six cross-sections (6 EU countries: Germany, Spain, France, UK, Italy, the Netherlands) had basically complete time series.[26]

(1a) Chile's market share in the EU with respect to the EU countries: sheu

$sheu_{ist}$ measures the share of Chilean exports (x) of sector s in EU country i at time t when competing against imports (m) from EU countries only:

$Sheu_{ist} = x_{ist}/m_{EUist}$

(1b) Chile's market share in the EU with respect to the non-EU countries: shnoneu

$shnoneu_{ist}$ measures the share of Chilean exports of sector s in EU country i at time t when competing against imports (m) from non-EU countries only:

$shnoneu_{ist} = x_{ist}/m_{non\text{-}EUist}$

(1c) Chile's market share in the EU with respect to the world (EU and non-EU countries): shw

shw_{ist} measures the share of Chilean exports of sector s in EU country i at time t when competing against imports (m) from EU and non-EU countries:

$shw_{ist} = x_{ist}/m_{EU+non\text{-}EUjst}$

(2) The Chilean real effective exchange rate: reer

[26] Due to missing data, Austria, Belgium, Finland, Luxemburg and Sweden were excluded from the analysis.

reer is the bilateral real effective exchange rate between Chile and the EU countries (price quotation system) from Chile's point of view. It consists of the real exchange rate (rer) and basic indicators of EU protection such as EU tariffs (t) and EU subsidies (s).

It is computed (all data for 'rer' are taken from World Development Indicators CD ROM of 2005) as:

$rer = e \cdot P_{EU}/P_{Chile}$ with

rer = real bilateral exchange rate between Chile and relevant EU country

e = nominal exchange rate (x Chilean Peso/1EUR) between Chile and relevant EU country

P_{EU} = GDP deflator of the EU country under consideration with 1995 as base year (1995 ≙ 100)

P_{Chile} = GDP deflator of Chile with 1995 as base year (1995 ≙ 100)

rer has been adjusted for EU tariff protection (in terms of average EU tariff rate (t)) and non-tariff protection (in terms of EU subsidy rate (s). Tariff rates prevailing in the EU can be found in Trade Policy Review European Union, Volume 1, 2000, pp. 88-101 (WTO) and rough subsidy equivalents are based on qualitative information on non-tariff protection collected, explained and nicely put together for UNCTAD by Supper (2001).

So we get:

$reer = rer \cdot (1-s)/(1+t)$

For the simulations, we assume that the FTA between Chile and the EU brings tariffs down to zero.

(3) Chile's competitors' (*) real effective exchange rates: reer*

In analogy to (2), we compute the real effective exchange rates of Chile's main competitors: Norway, Australia, South Africa, and Brazil. Nominal exchange rates, and Norway's, Australia's, South Africa's, and Brazil's GDP deflators are computed from World

Development Indicators CD ROM 2005. Tariff and subsidy rates are borrowed from WTO and UNCTAD (see (2)).

(4) Chile's copper price in real terms: rpcopper

$rpcopper = pcopper \cdot e_{RCHUS}/GDPDEFL_{RCH}$

with

pcopper = world market price of copper in US$ per ton

e_{RCHUS} = nominal exchange rate Chilean Peso/US$ (price quotation system)

$GDPDEFL_{RCH}$ = Chilean GDP deflator

Appendix 2

Figure 1: Chile's market share in EU fish imports with respect to EU and non-EU competitors in the period 1988–2002

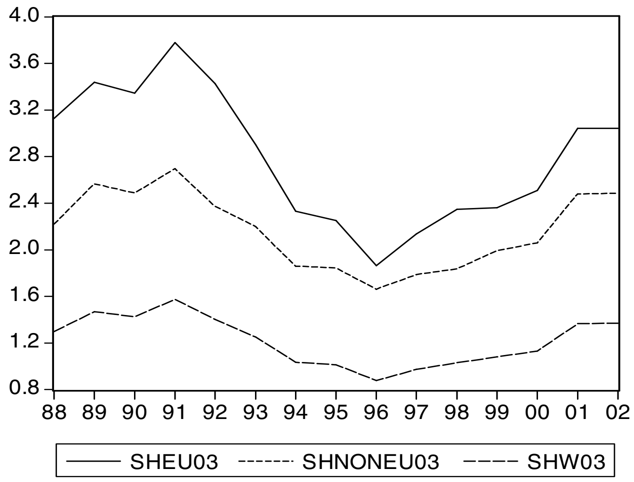


Figure 2: Chile's market share in EU fruit imports with respect to EU and non-EU competitors in the period 1988–2002

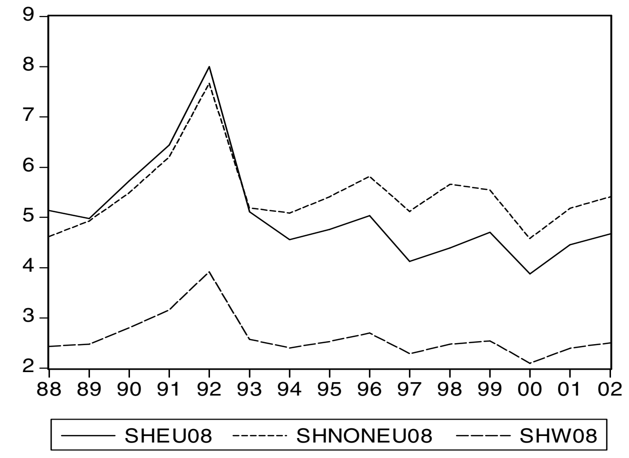


Figure 3: Chile's market share in EU imports of beverages with respect to EU and non-EU competitors in the period 1988–2002

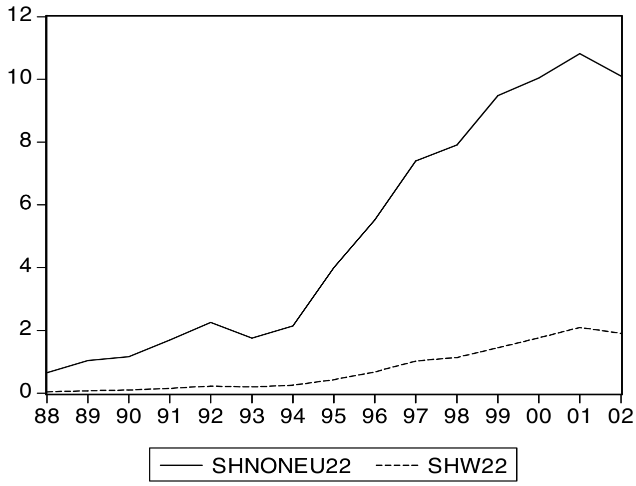


Figure 4: Chile's market share in EU imports of ores, slag, and ash with respect to EU and non-EU competitors in the period 1988–2002

Figure 5: Chile's market share in EU imports of wood and products thereof (44) with respect to EU and non-EU competitors in the period 1988–2002

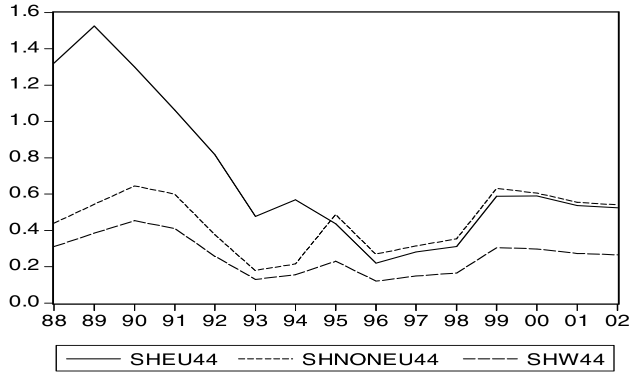


Figure 6: Chile's market share in EU's imports of wood pulp (47) with respect to non-EU and worldwide competitors in the period 1988–2002

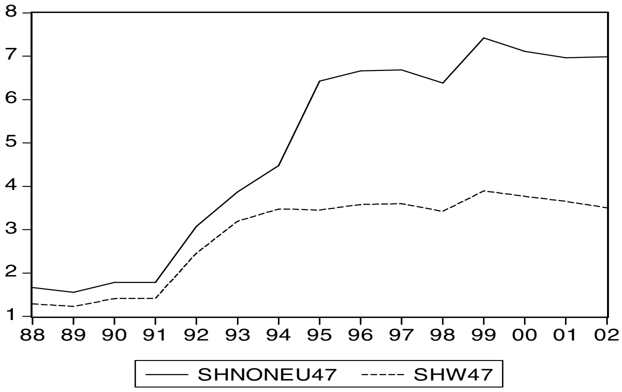
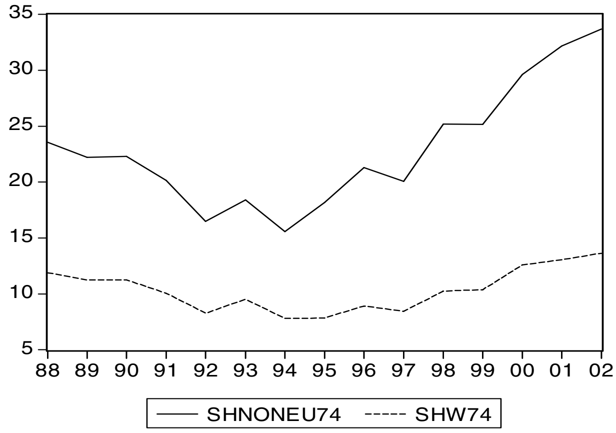


Figure 7: Chile's market share in EU's imports of copper (74) with respect to non-EU and worldwide competitors in the period 1988–2002



Modelling the Dynamics of Market Shares in a Pooled Data Setting: Econometric and Empirical Issues

Felicitas Nowak-Lehmann D.[a*], Dierk Herzer[b], Sebastian Vollmer[a], and Inmaculada Martínez-Zarzoso[c]

[a]*Ibero-America Institute for Economic Research at the University of Goettingen, Goettingen, Germany*

[b]*Goethe-University of Frankfurt, Frankfurt am Main, Germany*

[c]*University of Jaume I, Castellón, Spain*

The objective of this paper is twofold. First, it is to study the applicability of the widely used Autoregressive Distributed Lag Model (ARDL) in a pooled data setting. Second, it is to analyse Chile's market shares in the EU during the period 1988 to 2002, pointing to application problems that might jeopardise the model and searching for estimation methods that deal with the problem of inter-temporal and cross-sectional correlation of the disturbances. To estimate the coefficients of the ARDL model, Feasible Generalised Least Squares (FGLS) is utilised within the Three-Stage Least Squares (3SLS) and the non-standard Generalised Method of Moments (GMM) frameworks. A computation of errors is added to highlight the susceptibility of the model to problems related to the underlying model assumptions.

Keywords: dynamic panel data model; standard autoregressive distributed lag model; pooled Three-Stage Feasible Generalised Least Squares estimation; non-standard panel GMM estimation; market shares

JEL: F14; F17; C23

[*] Corresponding author. E-mail: fnowak@uni-goettingen.de

I. Introduction

In this paper, a standard Autoregressive Distributed Lag Model (ARDL) is utilised to estimate the dynamics of Chile's market shares in the EU market. This dynamic model has been adapted from studies of Balestra and Nerlove (1966), Baltagi and Levin (1986), Arellano and Bond (1991), Blundell et al. (1992), Islam (1995), and Ziliak (1997), Kim et al. (2003) among others. Cable (1997) applied an ARDL to market share behaviour and mobility in the UK daily newspaper market. A common feature of all these studies (and many more of this kind) is that the dynamic relationship between dependent and independent variables is captured by a lagged dependent variable, thus leading to an autoregressive distributed lag model. This is the standard dynamic model that is applied to panel data, as described in Baltagi (2005).

The main aim of this paper is to examine the applicability of the ARDL from both a theoretical and an empirical point of view. From a theoretical perspective, we analyse the structure and origin of this widely used autoregressive distributed lag model. From an empirical perspective, we illustrate estimation problems of the ARDL with an empirical application to Chile's market shares in the EU market. We differentiate among three types of caveats that can be lodged. The first is theoretical and deals with the underlying assumptions of the ARDL and the underlying geometric lag structure. The second caveat deals with the time-series properties of the data and the autocorrelation problem present in most panel data sets. The third caveat centres around the endogeneity of the lagged dependent variable on the right-hand side and the endogeneity of standard instrumental variables in the presence of serial autocorrelation. To tackle these estimation problems, the dynamic pooled data model is estimated by both the Three Stage Least Squares (3SLS) and non-standard Generalised Method of Moments (GMM) methods, in combination with Feasible Generalised Least Squares

(FGLS) and Seemingly Unrelated Regression (SUR) to deal with the problems of endogeneity and autocorrelation of the residuals across countries and over time.

Critically examining the preconditions of the model, studying its applicability to panel data, and highlighting the inherent problems of the ARDL are the main tasks of this paper. We differ from other dynamic panel studies in that we take into account the time-series properties of the variables, employ non-standard estimation techniques, and conduct an error analysis. To our knowledge, this is the first study to apply such procedures.

The paper is set up as follows. In Section II, applicability issues and estimation problems of the ARDL are discussed. Moreover, the derivation of the model and the underlying assumptions of the ARDL are analysed and combined estimation techniques to solve at least some of the estimation problems arising in a pooled data setting are then proposed. In Section III, we set up a simple dynamic market-share model for Chilean exporters to the EU and study the empirical applicability of the ARDL. Section IV presents our estimation results and an error analysis. Section V concludes.

II. The ARDL model in a panel/pooled data setting

The most widely used dynamic model for panel data is the first-order autoregressive distributed lag model with only a lagged dependent variable capturing the impact of current and lagged explanatory variables. For simplicity this will be called a simple autoregressive distributed lag model[1] (ARDL). The Panel ARDL (see Baltagi, 2005) is of the following form:

$$y_{it} = a + b_0 x_{1it} + c_0 x_{2it} + \ldots + q_0 x_{pit} + \lambda y_{it-1} + u_{it}\,, \tag{1}$$

[1] This is identical with the geometric lag model. The more complicated type of autoregressive distributed lag models corresponds to the transfer function model, also known as ARMAX model (for a good description, see Greene, 2000)

where i = 1, … , N, t = 1, … , T, a is a common intercept, $b_0, c_0, \ldots, q_0$ are the impact multipliers, λ is the adjustment parameter, and $\mu_{it} = \mu_i + \upsilon_{it}$. The two components of the error term are independent of each other and among themselves so that μ_i ~IID(0; σ_μ^2) and υ_{it} ~ IID(0; σ_υ^2) hold. This implies that autocorrelation of υ_{it} is assumed away. The simple ARDL model has become the most popular of all the dynamic models given that the lagged reaction between dependent and explanatory variables is captured in a single parameter, which is known as the adjustment parameter λ. This parameter expresses the reaction between y_{it} and $y_{it\text{-}1}$ explicitly and the reaction between y_{it} and x_{1it}, … , $x_{1it\text{-}k}$, …. , x_{pit}, … , $x_{pit\text{-}k}$ implicitly. λ 'summarises' the impact of all p-independent variables. In a panel or pooled data context, the ARDL can exist in two forms: the random-effects (RE) form and the fixed-effects (FE) form.

The advantage of the ARDL is that the number of lag coefficients to be estimated reduces to b_0,.., q_0 and λ and that all impact-coefficients b_1, q_k can be easily computed according to the general formula: $b_{\text{lag}} = b_0 \lambda^{lag}$, thus generating: $b_1 = b_0 \lambda^1$, $b_2 = \text{b}_0 \lambda^2$. A further advantage of the ARDL is that it is linear in its coefficients, thus allowing the application of linear estimation techniques.

To estimate the panel ARDL[2], various GMM estimators based on first differencing were developed by Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998), and Keane and Runkle (1992). These methods only ensure unbiased and consistent estimates if the υ_{it} are not auto-correlated.

Besides, the ARDL can be accompanied by two categories of problems that have not been dealt with in the dynamic panel literature. The first is related to the underlying model and its very restrictive assumptions. The second has to do with the time-series properties of the variables and the high likelihood of autocorrelation of the disturbances that renders the application of standard GMM techniques inappropriate.[3] We discuss these problems in the following two subsections.

The underlying model, the Koyck transformation, and the ARDL assumptions

Koyck (1954) was the first to show how an ARDL model (Equation 1) can be obtained from a geometric lag model:

$$y_{it} = a_i + b_0\lambda^0 x_{1it} + \ldots + b_0\lambda^k x_{1it-k} + \ldots + q_0\lambda^0 x_{pit} + \ldots + q_0\lambda^k x_{pit-k} + \upsilon_{it}, \tag{2}$$

where $0 < \lambda < 1$, $b_k = b_0\,\lambda^k$ denotes the impact of a change that happened k periods ago, and λ is the same for all regressors x_{1it}, … , $x_{pit.}$ This transformation is called Koyck (lag) transformation. Equation 2 lagged by one period and multiplied through with λ gives:

$$\begin{aligned} \lambda y_{it-1} = \lambda a_i + b_0\lambda^0 \lambda x_{1it-1} + \ldots + b_0\lambda^k \lambda x_{1it-k} + \ldots + \\ q_0\lambda^0 \lambda x_{pit-1} + \ldots + q_0\lambda^k \lambda x_{pit-k-1} + \lambda\upsilon_{it-1} \end{aligned} \tag{3}$$

By subtracting Equation 3 from Equation 2 and *ignoring* the terms $b_0\,\lambda^{k+1} x_{1it-k-1}$, … , $q_0\,\lambda^{k+1} x_{pit-k-1}$, we eventually obtain the ARDL (Koyck lag formulation of the geo lag model):

$$y_{it} = a_i^* + b_0 x_{1it} + c_0 x_{2it} + \ldots + q_0 x_{pit} + \lambda y_{it-1} + v_{it} \tag{4}$$

where $a_i^* = a_i(1-\lambda)$ and $a_i = a_i^*/(1-\lambda)$ and $v_{it} = \upsilon_{it} - \lambda\, \upsilon_{it-1}$.

In the geometric lag model and its Koyck transformation (the ARDL), all explanatory variables (x_{1it}, x_{2it}, … , x_{pit}) have a geometrically declining impact on the

---

[2] For efficient estimation of models for panel data see Ahn and Schmidt (1995) and Anderson and Hsiao (1982).

dependent variable y_{it}, in such a way that changes in the distant past have a more minor impact than changes in the more recent past (see Figure 1). In a multivariate dynamic regression model, all explanatory variables (x_{1it}, x_{2it}, … , x_{pit}) have to impact on y_{it} in exactly the same geometric way, with the same λ. This pre-condition can become extremely restrictive.[4]

[Figure 1 about here]

In addition, there are many instances in which the assumption of a geometric lag itself will not be fulfilled. This will be true especially when reaction lags are present and when changes in the current and the preceding periods therefore have a lesser impact than changes in earlier periods. In such cases, a better option is a polynomial lag model, which allows us to estimate any lag structure that can be depicted by a polynomial of order 1, 2, … , p.

A severe shortcoming of the ARDL is that the estimators will be inefficient and biased (even inconsistent) in the presence of autocorrelation of the disturbances (Kelejian and Oates, 1981). Additionally, the problem of neglecting $b_0 \lambda^{k+1} x_{1it-k-1}$ can lead to huge errors (estimation mistakes) if λ is relatively large and the maximum lag, k, is short. A short lag length might be a problem when working with annual data and less of a concern when working with daily or weekly data. This point will be elaborated and illustrated in the empirical results section (Section IV, Table 3).

Estimation problems arising in an ARDL with longitudinal (pooled) data

Dynamic panel data models usually use a small number of observations over time, whereas pooled data models are based on longer time spans (the number of

[3] Standard GMM utilises lagged variables as instruments. This leads to biased estimates in the presence of autocorrelation of the disturbances.
[4] In a bi-variate regression model, this assumption may not be so restrictive, but the assumption will also not always be fulfilled.

observations over time is usually larger than the number of cross-sections, $T > N$). However, establishing a dynamic structure in the form of an ARDL with only a small number of observations over time (e.g., $T = 2$ or $T = 3$) is not a satisfying approach since building a dynamic model based on a very small number of observations over time can only capture the true dynamics "by chance". If possible, one should work with a longer time span (pooled data setting).

In settings with larger T, the time-series properties of the variables become relevant. All the time-series problems must be dealt with and the time-series properties of the variables must be scrutinised very carefully in order to avoid running spurious regressions.[5] Scrutinisation of the series implies checking whether the series of the regression model have a memory, that is to say, whether y_{it}, x_{1it}, …, x_{pit} are determined by their past values. As is well known from the time-series literature, the unit root tests[6] check whether a series (see equations 5 through 7) is non-stationary, whether it has a unit root, with $|\rho_i| \geq 1$ in the H_0-hypothesis (Davidson and MacKinnon, 1993, Hamilton, 1994, and Hayashi, 2000).[7]

$$y_{it} = \rho_{yi} y_{it-1} + u_{it} \tag{5}$$

$$x_{1it} = \rho_{x1i} x_{1it-1} + u_{it} \tag{6}$$

$$x_{pit} = \rho_{xpi} x_{pit-1} + u_{it} \tag{7}$$

[5] This problem was discussed extensively by Granger and Newbold (1974) and led to the development of stationarity tests (unit-root tests) such as the Augmented Dickey-Fuller test (1979), the Phillips-Perron test (1998), the Kwiatkowski, Phillips, Schmidt, and Shin test (KPSS, 1992), the GLS-detrended Dickey-Fuller test (Elliott, Rothenberg, and Stock, 1996), the Elliott, Rothenberg, and Stock Point Optimal test (ERS, 1996), and the Ng and Perron test (NP, 2001).

[6] In the last ten years, enormous progress has been made in the field of panel/pool unit root tests. There are two types of panel/pool unit root tests. One type assumes panel homogeneity (common unit root processes for all cross-sections). These tests were developed by Levin, Lin, and Chu (2002), Breitung (2000), and Hadri (1999). The second type of tests allows for panel heterogeneity. The Im-Pesaran-Shin test (IPS test, 2003), the Fisher-type tests using ADF, and the Phillips-Perron test (Maddala and Wu, 1999; Hadri (2000), Choi (2001), Hadri and Larsson (2005)) are based on individual unit roots or coefficients ρ_i for each cross-section.

[7] For simplicity, neither constants nor trends are included in the equations.

If the series in the regression model has a strong memory of the past, there is a high likelihood that the omitted variables, which are lumped together in the error term v_{it}, will have a strong memory of the past, as well. They need not necessarily possess a unit root (be non-stationary with a ρ close to one), but the probability that the error terms will be auto-correlated is high, i.e., $v_{it} = \rho_i v_{it-1}$, with ρ_i being significantly different from zero.

Hujer et. al. (2005), Kim et al. (2003), Sevestre and Trognon (1996) and Keane and Runkle (1992) have studied the issue of serial correlation of the disturbances. To tackle serial correlation, Keane and Runkle (1992) and Kim et al. (2003) propose the forward-filtering 2SLS method (KR estimate). This method pretends serial correlation to be equal to 1 ($\rho_{ik} = 1$), which is a very rough estimate. Kim et al. (2003) refine the KR method. We, in contrast, estimate the extent of serial correlation in the sample (our $\hat{\rho}_{ik}$)[8] and then transform the variables correspondingly (in soft or quasi-first differences) applying the FGLS technique.[9]

However, autocorrelation of the residuals is not the only problem. When a lagged endogenous variable appears on the right-hand side of a regression equation (as in Equation 4) and when the disturbances are autocorrelated, the lagged endogenous variable will automatically be correlated with the disturbance term and thus become endogenous. The endogeneity problem of the lagged dependent variable (y_{it-1}), which is caused by first-order AR correlation of the residuals due to non-stationarity of the series, can be effectively tackled by the Two-Stage Least Squares technique utilising 2SFGLS. Typical standard GMM procedures, in contrast, such as those summarised in

[8] In FGLS, the unknown serial correlation coefficient is estimated as described in Section II.
[9] In samples with sufficiently large T[9] and errors that follow an AR process, ECM or Dynamic Ordinary Least Squares (DOLS) techniques can be applied, but in samples with shorter T, FGLS techniques are preferable.

Baltagi (2005) cannot be applied in the presence of autocorrelation since the instruments (variables with two or more lags; both in levels or in differences) used for the lagged dependent variable will also be correlated with the error term through υ_{it-1}.

Moreover, cross-section correlation among the residuals is expected to be a very probable feature in pooled data sets. In this instance, it is advisable to build a system of equations (with one equation per cross-section) and estimate the system with the Seemingly Unrelated Regression (SUR) technique[10]. If the SUR-technique is combined with Two-Stage Least Squares (to control for endogeneity) it is called Three-Stage Least Squares (3SLS). If it is furthermore combined with FGLS (to control for autocorrelation), we will call it 3SFGLS. This technique, as well as non-standard GMM combined with SUR (to control for cross-section correlation), will be applied in Section IV.

III. The model specification and estimation techniques

From an applied economist's point of view, the objective of this paper is to analyse Chile's market share in the EU market on a sectoral level over the period from 1988 to 2002, applying the necessary panel/pooled time-series techniques. The ARDL model is built on six cross-sections (EU countries) and 15 annual observations for Chile's seven most important export sectors (fish, fruit, wine, ores, wood, wood pulp, and copper).

The market share ARDL model

Following Sutton (2004), there are two contradicting views on the development of market shares over time: the first goes back to Alfred Chandler and asserts that

[10] Building a system is only possible if the number of cross-sections is small.

market shares are robust over time and that leadership tends to persist for a substantial time. The second view, propagated by Schumpeter, emphasises the transience of leadership positions. Schumpeter labels those leadership positions that arise from invention and innovation *temporary monopolies*. However, there is no benchmark for long or short leadership positions (2002 Japan Conference, 2005).

We will test the relevance of these hypotheses by means of panel/pooled unit root tests. If market shares turn out to be stationary (I(0)), this will indicate that they are robust and persistent during the period from 1988 to 2002. However, if they are instead non-stationary, then we will conclude that the Schumpeter hypothesis cannot be rejected by the data. Of course, our time period is too short to draw conclusions about whether the Schumpeter hypothesis is valid in the long run.

There are also two econometric approaches to modelling market shares: According to the first, market shares are basically purely stochastic, and according to the second, they are influenced by hard economic factors such as prices, marketing expenditures, the number and strength of competitors, etc. To model market shares, Sutton (2004) chooses a mix of the first and the second approaches (eclectic approach). Favouring the idea of building a stochastic model, he expands the model to include industry-specific features (e.g., a strategic representation of firms' competitive responses to market-share changes). However, it has to be kept in mind that strategic behaviour is very often intrinsically unobservable.

Cable (1997) models market shares according to the eclectic econometric approach and uses an autoregressive distributive lag model (ARDL). He selects a first-order autoregressive model with a one-period lagged endogenous variable implying a temporary persistence of market shares. In his model prices and advertising shares[11] are

[11] Advertising is important when selling a differentiated product, but not when selling rather homogeneous products (as in our case study).

the explanatory variables for the UK's national daily newspapers. Other authors also emphasise the importance of non-price factors as explanatory variables for the export market share (stemming from advertising and technological advantage (research and development)) when studying industrialised countries and the manufacturing sector (Hula, 1989; Das et al., 1993; Amable and Verspagen, 1995).[12]

We follow Cable's approach in terms of dynamic modelling, but not in terms of the determining variables. We stress the role played by observable and quantifiable factors, such as bilateral real effective exchange rates[13]. Thus we believe that exchange rates, cost differentials, tariffs, and subsidies are important *hard* factors explaining market shares over time. Accordingly, we build a dynamic econometric model in which price competitiveness (Chile's and its competitor's bilateral real effective exchange rate) is considered decisive for the competitive position (see Equation 8). Price competitiveness is considered a decisive determinant of Chile's market shares, since Chile's most successful exports are rather homogeneous products (fish, fruit, beverages, ores, copper, and wood and products thereof). Thus, our empirical model is of the following form:

$$lshw_{ist} = a_{is} + \beta_0 lreer_{ist} + \gamma_0 lreer^*_{ist} + \lambda lshw_{ist-1} + v_{ist} \tag{8}$$

$$\begin{aligned} lshw_{ist} = a_{is} + \beta_0 lreer_{ist} + \beta_0 \lambda^{k+1} lreer_{ist-k-1} + \\ \gamma_0 lreer^*_{ist} + \gamma_0 \lambda^{k+1} lreer^*_{ist-k-1} + \lambda lshw_{ist-1} + v_{ist} \end{aligned} \tag{9}$$

Equation 8 is the simple ARDL and Equation 9 is the complete Koyck transformation, where $i = 1, 2, \ldots, 6$ represents the cross-sections: France (FRA), the Netherlands (NDL), Germany (DEU), Italy (ITA), Great Britain (GBR), and Spain (ESP); t = 1988,

[12] We fully agree with the importance of non-price factors in the industry sectors.
[13] The bilateral real exchange rate captures the depreciation of the euro vis-á-vis the US dollar in the 1995 to 2001 period and the appreciation of the Chilean peso with respect to the US dollar in the period from

1989, … , 2002 are years (annual observations), s = 03, 08, 22, 26, 44, 47, and 74 are the sectors (according to the two-digit HS classification), $lshw_{ist}$ stands for Chile's market share in EU country i in sector s at point t, $lreer_{ist}$ is Chile's real effective exchange rate, prevailing in EU country i and in sector s, and $lreer*_{ist}$ is Chile's competitor's (*) real effective exchange rate, prevailing in country i and in sector s. Since *lreer* and *lreer** are in price quotation, we expect *lreer* to have a positive impact and *lreer** to have a negative impact.

Market shares in a specific sector (s) are computed as a ratio of Chile's sectoral exports (X in the numerator) and EU country i's imports from the world $M_{.i} = M_{\mathrm{EU}i} + M_{\text{non-EU}i}$ (in the denominator). Due to unsubstantial trade volumes, we consider only Chile's market shares in France, the Netherlands, Germany, Italy, UK, and Spain. Market shares are computed for seven sectors at the two-digit HS levels, namely, fish (03), fruit (08), beverages (22), ores (26), wood (44), wood pulp (47), and copper (74). The data are from COMEXT and TradeCAN.

As to the peso-euro development, the peso appreciated in general terms against the euro from 1988 to 2002 so that Chile's price competitiveness was dampened.[14] This effect was due to appreciation of the peso against the US dollar from 1993 to 1997 and depreciation of the euro (or of the DM, French franc, lira, etc.) in relation to the US dollar over the period from 1995 to 2001. However, this effect is captured in the bilateral real exchange rate and does not require the introduction of a dummy variable.

A detailed description of the data (including their calculation) can be found in the appendix of the working paper version of this paper or at http://www2.vwl.wiso.uni-goettingen.de/iberoAppendix_AECON_08.pdf. The estimation period is from 1988 to

1993 to 1997. In the period from 1988 to 2002, Chile's price competitiveness was by and large impeded by the appreciation of its currency vis-á-vis the euro.

2002. Thus, we obtain a maximum of six cross-sections and 15 years, resulting in a maximum of 90 observations per sector. The number of observations varies depending on the sector studied.

We estimate Equation 8 as a fixed-effect model allowing for cross-section-specific intercepts. This model could be applied in its unrestricted form by estimating cross-section-specific slope parameters for $lreer_{ist}$, $lreer*_{ist,}$ and $lshw_{ist-1}$ (β_{0i}, γ_{0i}, and λ_i), but given our limited number of observations in each cross-section, we stick to common slope parameters in all countries.

As shown in Section II, Equations 8 and 9 are derived from the geometric lag model:

$$\begin{aligned} lshw_{ist} = a_{is} + \beta_0 \lambda^0 lreer_{ist} + \ldots + \beta_0 \lambda^k lreer_{ist-k} + \\ \gamma_0 \lambda^0 lreer*_{ist} + \ldots + \gamma_0 \lambda^k lreer*_{ist-1} + v_{ist} \end{aligned} \tag{10}$$

As to the coefficients and the disturbance in this type of model, it is assumed that $0 < \lambda < 1$ and that λ is the same for all regressors. Furthermore, it is assumed that $\beta_{lag} = \beta_0\, \lambda^{lag}$, $\gamma_{lag} = \gamma_0\, \lambda^{lag}$, and $v_{ist} \sim N(0; \sigma_v^2)$. Note that Equation 10 assumes not only a geometric reaction of the market share (*lshw*) with respect to relative prices (β_i and γ_i must follow a geometric lag) in all six importing countries i under investigation, but it assumes exactly the same geometric reaction (as measured by λ_i) of $lshw_{ist}$ with respect to changes of all the regressors (both $lreer_{ist}$ and $lreer_{ist}*$). In our case, as well as in many other studies using the ARDL, the above assumption cannot be justified by the data for all regressors. Also, the specific geometric reaction does not always apply to all countries under study. These issues would become even more

[14] Appreciation of the peso was less pronounced with respect to the British pound (GBP).

crucial with an increasing number of cross-sections and with some more explanatory variables in the model (a model with, for example, 100 countries and five regressors).

Therefore, before applying our data to the ARDL, we examined the cross-correlations between the dependent and the independent variables[15] (12 per sector, 84 cross-correlations in total). With the help of cross-correlations, the dynamics of the model (the lag structure between dependent and independent variable) can be studied. The cross-correlations indicate that the geometric lag assumption is not fulfilled in a variety of cases and that the maximum lag length is between two and three years.

Estimation techniques in the presence of non-stationary data

a) Testing the time-series properties of the data

In the first step, we test the time-series properties of the data (all in natural logs). All series, i.e., market shares (*lshw*), Chile's real effective exchange rate (*lreer*) and Chile's competitors' real effective exchange rates (*lreer**) for all country pairs are subjected to tests of non-stationarity (panel unit root tests). This procedure is applied to all seven sectors under investigation, neglecting the possible existence of structural breaks in the series because neither fundamental, abrupt changes in economic policy, nor any major exogenous shocks were detected in the period from 1988 to 2002.[16]

In the statistical analysis we allow for different unit root processes in the panel, i.e., cross-section-specific (country-specific) unit roots. We apply the Im, Pesaran, and Shin (2003) panel unit root test to all series considering the possibility of individual unit roots of our panel data. We find that almost all variables (*lshw*, *lreer*, and *lreer**) are

[15] These cross-correlations show the reaction pattern between the dependent and the independent variables very clearly and should precede the building of any dynamic models. The 84 cross-correlations are available from the authors upon request.
[16] The economic policy of the Pinochet government was continued under the governments of Aylwin, Frei, and Lagos. Consequently, the time series display no sign of a significant structural shift.

non-stationary, and integrated of order one, I(1) (results are not reported to save space). Of course, we have to be cautious in interpreting the results, since unit root tests generally tend to falsely accept the unit root null in small samples. Nevertheless, this result of non-stationarity is in line with our finding that in general the residual terms follow an AR(1) process (AR processes have a long memory[17]) and not an MA(1) process (MA processes have a short memory (Hujer, 2005)). Besides that, we can already conclude from the plots of the data (available at: http://www2.vwl.wiso.uni-goettingen.de/ibero/Appendix_AECON_08.pdf) that the market shares exhibit non-stationary behaviour.

With respect to market shares, this finding supports Schumpeter's view that gains in market shares are temporary. Monopolistic positions have to be defended; otherwise they are lost quickly. This view seems to apply especially to the fish, fruit, beverages, ores, and copper sectors. Market shares appeared more stable in the wood sectors (44 and 47) (see http://www2.vwl.wiso.uni-goettingen.de/ibero/Appendix_AECON_08.pdf), but are non-stationary according to the tests. This is in line with the results of Resende and Lima (2005), who found market share instability and market rivalry in the Brazilian industry utilising panel unit root tests, as well.

b) The FGLS approach versus co-integration approaches

When all variables are I(1), one could proceed with co-integration analysis and panel co-integration tests (Pedroni, 1999; Pedroni, 2004; Breitung and Pesaran, 2005). However, co-integration is a long-term concept that is not applicable to our rather short time span. Moreover, with fifteen annual observations, the power of panel co-

[17] Macroeconomic data usually show unit roots in the series and are therefore plausibly characterised by an autoregressive error process.

integration tests would be extremely low.[18] But co-integration analysis is not the only approach that deals with non-stationary series, and yields unbiased and efficient estimates in a dynamic model.[19] FGLS is another possibility, as is known from time series analysis. Therefore, we exploit the special suitability of FGLS for estimating dynamic models with panel data (see Stock and Watson, 2003).

In a panel/pooled analysis setting, FGLS works in analogy to the time series setting. The idea remains the same: non-stationarity of the series in a regression equation is reflected in the autocorrelation ρ of the residuals over time. Annual data usually shows first-order autocorrelation, and this is the case in our sample, as well.[20] The FGLS method is applied in three steps. First, Equation 8 is estimated by SUR and the residuals $\hat{v}_{it}$ are computed. Second, the order (first-order, second-order, or p-order) of autocorrelation $\hat{\rho}_k$ is estimated applying SUR and testing its significance[21] in Equation 11:

$$\hat{v}_{ist} = \sum_{k=1}^{K} \rho_{isk} \hat{v}_{ist-k} + e_{ist}, \tag{11}$$

where $e_{ist} \sim N(0; \sigma_{ei}^2)$ and $k = 1, 2,\ldots, K$ is the number of lags. Third, if only first-order autocorrelation is present (as in our case), the variables of Equation 8 are transformed into

$$lshwz_{ist} = lshw_{ist} - \hat{\rho}\, lshw_{ist-1}, \tag{12}$$

$$lreerz_{ist} = lreer_{ist} - \hat{\rho}\, lreer_{ist-1}, \tag{13}$$

[18] We have estimated the market share dynamics with an ECM based on an ARDL. However, using this procedure, we could explain much less of the variation of market shares, i.e., our R^2 (adjusted) were much smaller.
[19] Rao (2007) reviews three alternative approaches, viz., general to specific, vector autoregressions, and vector-error correction models, to estimate short and long-run relationships.
[20] ρ is usually well below 1; first differencing thus is a very imprecise (ineffective) method to remove stationarity.

$$lreerz*_{ist} = lreer*_{ist} - \hat{\rho}\, lreer*_{ist-1}, \quad (14)$$

$$lshwz_{ist-1} = lshw_{ist-1} - \hat{\rho}\, lshw_{ist-2}, \text{ and} \quad (15)$$

$$\varepsilon_{ist} = \hat{v}_{ist} - \hat{\rho}\, \hat{v}_{ist-1}, \quad (16)$$

thus generating variables in soft or quasi-first differences. Equation 8 is then estimated on the basis of the transformed variables (see Stock and Watson, 2003):

$$lshwz_{ist} = a_{is}(1-\hat{\rho}) + \beta_0 lreerz_{ist} + \gamma_0 lreerz*_{ist} + \lambda\, lshwz_{ist-1} + \varepsilon_{ist}. \quad (17)$$

IV. Estimating the market share ARDL

For each sector, separate panel ARDLs (applying Equation 8) are run over the time period 1988 to 2002, with the EU countries acting as cross-sections in the panel analysis. To control for autocorrelation, FGLS is combined with either 3SLS based on a system of equations or a non-standard GMM-type routine. Our system contains six equations, one for each cross-section/destination market/EU market[22]. Possible cross-equation/cross-section correlation of the error terms is controlled for by estimating the system of equations by means of SUR. Cross-section correlation can result from for example inter-related shifts in preferences (a shift in favour of biologically produced fish or fruit or in favour of higher quality wine, wood, copper, and ores etc.) that happens in all or some of the six EU markets under study. The presence of autocorrelation leads us to use instruments from outside the system, instead of using lagged values of the endogenous variables as instruments (which is the standard

[21] In our data first-order autocorrelation of the type $\hat{v}_{it} = \hat{\rho}_{i1}\ \hat{v}_{it-1}$ turns out to be present and dominant. $\hat{\rho}_{i1}$ expresses first-order autocorrelation, henceforth to be called $\hat{\rho}$.

[22] The system/SUR approach is recommended when the number of N is small (six in our case) and T is large (15 in our case).

technique). These instruments are utilised both in the 3SLS (2SLS-SUR) routine and in the non-standard GMM routine.

The 3SLS approach

The choice of instruments is crucial in order to obtain consistent estimates in any model, including in the market share model. We used an indicator of production capacity in real terms as an instrument for lagged market share ($lshw_{ist-1}$), the difference in PPP-income between Chile and the importing country as an instrument for $lreer_{ist}$, and the competitor's real exchange rate in a transformation that is generally used in polynomial lag models as an instrument for $lreer^*_{ist}$. Table 1 summarises the impact of price competitiveness on market shares estimated by Three-Stage Feasible Generalised Least Squares (3-SFGLS).

[Table 1 about here]

Under the assumption that the data follow an ARDL model, we find a significant positive impact of increased Chilean price competition on market shares in the fish (03), fruit (08), and ores (26) sectors, but no significant negative impact of *foreign* price competition on market shares in six out of seven sectors under study. As to beverages (22) which mainly consist of wine exports, we find a negative impact of competitive (low) Chilean prices and a positive impact of low foreign prices on market shares. FAO statistics (FAO Production Yearbook, 2003; FAO Trade Yearbook, 2003) show that Chile increased its wine production in the period 1978 to 2002. Such a production increase, which is usually achieved by intensified irrigation and fertilisation, leads to inferior wines at lower prices. Our regression results indicate that consumers in the EU associate low price with low quality and therefore reduce demand. Therefore, we tend to conclude that quality considerations dominate price considerations in the beverages sector. Adjustment to the long-run equilibrium was significant in the beverages (22), ores (26), wood (44), wood pulp (47), and copper (74) sectors, whereas no significant adjustment took place in the fish (03) and fruit (08) sectors.

The non-standard GMM-type approach

In the absence of serially correlated error terms, the standard (classic) GMM approach does have a comparative advantage over 3SLS in controlling endogeneity. Control of endogeneity is 100 percent due to specific model restrictions and therefore leads to a gain in unbiasedness. However, efficiency is lost by creating a tremendous amount of moment conditions that have to be taken into account. In our case, we get 210 moment conditions, i.e., 210 restrictions[23], which fact highlights the computational burden of this approach (Schmidt et al., 1992).

However, in the presence of autocorrelation of the disturbances (as in our case study), the standard GMM approach, which uses lagged variables as instruments for endogenous regressors, must be avoided, since this strategy of creating instruments for endogenous variables fails due to autocorrelated errors (Durlauf et al., 2004). This shortcoming also applies to the Arellano-Bond (1991) estimator which is based on running the regression in first differences.

In order to solve the problem of endogeneity, we estimate the dynamic model by non-standard GMM (for GMM see: Holtz-Eakin et al., 1988; Arellano and Bond, 1991; Caselli, Esquivel and Lefort, 1996; Durlauf et al., 2004) without utilising lagged variables as instruments. Instead, we take exactly the same ones as in the 3SLS routine

[23] The number of restrictions is $T(T\text{-}1)\,K/2$.

described in the previous section: the difference in PPP income between Chile and the importing country, an indicator of production capacity in real terms, and the real exchange rate in a transformation that is generally used in polynomial lag models.

[Table 2 about here]

Assuming for the moment that the underlying preconditions of the autoregressive lag model are fulfilled, we can conclude from Table 2 that there is a positive relationship between an increase in Chilean price competitiveness and market share in the fruit sector (08) and a negative relationship between low Chilean wine prices (sector 22) and high Chilean copper prices (sector 74) and their respective market shares. Foreign relative prices have a significant impact in the fish (03), beverages (22), and ores (26) sectors. The role of prices in the wood (44) and wood pulp (47) sectors might be severely impeded by illegal logging and illegal imports of wood products. Illegal logging distorted official trade flows not only of all timber products (roundwood, sawn wood, veneer, plywood, boards, semi-finished and finished products, and furniture), but also of pulp, paper, printed products, and cellulose.[24] This latter statement applies also to the interpretation of the 3SLS estimation.

Error analysis

The results concerning the slope coefficients of the price competitiveness variables must be taken with caution if the actual lag length of market-share dynamics is small. In our case study, the maximum lag length was about two to three years

[24] Illegal logging is estimated to comprise up to 50 percent of all logging activity in the key countries of Eastern Europe and Russia, up to 94 percent in the key Asian countries, up to 80 percent in the key African countries, and up to 80 percent in the key Latin American countries (WWF, 2005; FERN, 2004).

according to the cross-correlations. When transforming Equation 10 into the simple ARDL (Equation 8) (which must be considered an incomplete Koyck transformation), an error occurs by suppressing the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer*_{ist-k-1}$. The shorter the actual lag ($k_{\max}$) and the closer λ (the adjustment parameter) is to 1, the larger this error is. The extent of the error becomes intuitively clear by treating Equation 9 as the true ARDL and considering it as the complete Koyck transformation.

The actual error computation is very straightforward. If the maximum actual lag is k, then the error occurs by dropping the terms $\beta_0 \lambda^{k+1} lreer_{ist-k-1}$ and $\gamma_0 \lambda^{k+1} lreer*_{ist-k-1}$ is λ^{k+1}. This implies that a maximum lag length of one (two) will lead to an error of λ^2 (λ^3). When working with annual data, a specification with one or two- year (maximum) lags can be very common, such that the danger of committing an error is relatively high.

[Table 3 about here]

We can draw several conclusions from the error analysis in Table 3:

(1) The data do not fit the autoregressive lag model in the fruit sector (3SLS and GMM estimation) or in the fish sector (3SLS estimation). The λs there carry the wrong sign and are insignificant since the ARDL requires significant positive λs that lie in an interval $]0;1[$.

(2) The data can be explained by an ARDL in the rest of the sectors by and large since the λs lie in an interval $]0;1[$. However, since we work with annual data where the maximum lag length is usually short ($k_{\max}$ = 2 is very realistic according to the cross-correlations), large errors will result in the beverages, ores, and copper sectors; λ is relatively large and the omission of the terms

$\beta_0\, \lambda^{k+1}$ lreer and $\gamma_0\, \lambda^{k+1}$ lreer* will therefore result in a large error. For example, for $\lambda = 0.80$, the error is 64 percent if $k_{\max}$ is 1, and 51 percent if $k_{\max}$ is 2. That is, 64 percent or 51 percent of the impact of copper prices on the market share in copper is neglected. Large errors also occur in the beverages, ores, and wood sectors given that λ is relatively large there.

(3) Note that the errors are even larger than computed when we have reason to assume that the geometric lag structure does not apply in all instances. Computation of errors in this case would require knowledge of the true model.

Comparison of the 3SLS and the non-standard GMM results

On the one hand, we have found that the ARDL estimations in Section IV have very respectable adjusted R^2 measures and Durbin-Watson (DW) statistics around 2.[25] On the other hand, the standard errors of the regressions are relatively high. Moreover, the error analysis makes clear that the simple dynamic specification in the form of an ARDL suffers from some drawbacks. The autoregressive lag specification does not seem to apply in the fish or the fruit sectors. Statements in the beverages, ores, wood, and copper sectors are subject to relatively large errors due to neglecting the term λ^{k+1}, the impact of changes in prices, and protection[26] in the autoregressive transformation.

The estimation results of 3SLS and non-standard GMM differ widely. This result is puzzling since fixed effects and exactly the same instrumental variables are utilised in both estimation procedures. However, 3SLS and non-standard GMM differ in the number of restrictions applied. 3SLS basically works under the condition of

[25] Even though the DW statistic must be adjusted in the presence of a lagged endogenous variable, the DW statistic is still able to roughly indicate problems of autocorrelation and misspecification. A better measure of autocorrelation is probably Bhargava's et al. (1981) DW statistic.

minimising the squared residuals of Equation 8 with instruments replacing the right-hand-side variables. GMM estimation, in contrast, is built around a multitude of moment conditions, some of which will be relevant and others irrelevant. The GMM routine does not involve a search for relevant moment conditions, and thus some irrelevant moment conditions can become binding (see Ziliak, 1997). Therefore, in our view, 3SLS is superior to GMM.

Moreover, it must be noted that the estimation results of both 3SLS and non-standard GMM do not fulfill our expectations as far as signs (especially in GMM) and significance are concerned. This certainly has to do with violated model assumptions but also with the simplicity of the model (we do not control for quality). Therefore, the empirical results should not be overemphasised, nor should they be utilised for further analysis.

V. Conclusions

Assuming that the underlying geometric lag specification can be applied to the data, the ARDL specification allows us to draw correct inferences about the short, medium, and long run. The ARDL specification can be combined with the FGLS and the SUR technique and is therefore able to deal with several estimation problems resulting from autocorrelation, heteroscedasticity and cross-sectional correlation of the disturbances. Applied to a system of equations, this technique transforms the variables in the regression equation by working with soft differences in the variables and by weighting the regressor matrix with a weight matrix that can control for heteroscedasticity of the variance of the residuals (White method) and for cross-sectional correlation of the

[26] All our prices contain sector-specific protection whenever relevant.

disturbances (SUR method). The endogeneity problem is solved with instrumental variables IV in either a 3SLS or a non-standard GMM-type routine. Unlagged exogenous variables are utilised to control for the endogeneity problem and to obtain unbiased estimates. Furthermore, the 3SFGLS and the non-standard GMM-type technique are able to produce efficient and consistent estimates if ARDL is the true model.

Violation of the geometric lag assumption is to be expected in particular when working with heterogenous panel data and with multivariate regression models, and will result in inconsistent estimators. In this case, a polynomial lag model could be the model of choice if there is not excessive cross-sectional heterogeneity. Estimations in the framework of panel error correction models and panel DOLS could be highly advisable even though these models require much longer time spans to allow for meaningful panel unit root and panel co-integration tests. Further research is needed on this topic.

Our study has demonstrated that the ARDL model must be applied with caution for several reasons. First, the geometric lag assumption was not supported overall by the cross-correlations between dependent and independent variables. Second, a maximum lag length of two to three years (also visible in the cross-correlations) can result in substantial estimation errors. Third, non-stationarity of the series leads in general to autocorrelation of the residuals. It renders the utilisation of lagged instruments in a standard GMM framework obsolete and requires a search for new instruments, which instruments, however, may not be applicable in all cases.

Acknowledgement

We are grateful to Stephan Klasen and an anonymous referee for excellent suggestions that considerably improved this paper.

References

Ahn, S. and Schmidt,P. (1995) Efficient estimation of models for dynamic panel data, *Journal of Econometrics* **68**: 5-27.

Amable, B. and Verspagen, B. (1995) The role of technology in market shares dynamics, *Applied Economics* **27**: 197-204.

Anderson, T.W. and Hsiao, C. (1982) Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* **18**: 47-82.

Arellano, M. (1989) A note on the Anderson-Hsiao estimator for panel data, *Economic Letters* **31**: 337-341.

Arellano, M. and Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* **58**: 277-297.

Balestra, P. and Nerlove, M. (1966) Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas, *Econometrica* **34**: 585-612.

Baltagi, B.H. and Levin, D. (1986) Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation, and advertising reconsidered, *Review of Economics and Statistics* **68**: 148-155.

Baltagi, B.H. (2005) *Econometric Analysis of Panel data*, Third Edition, Colchester: John Wiley & Sons, Ltd.

Bhargava, A. (1981) Serial correlation and the fixed effects model, *Working Paper 33*, London School of Economics and Political Science, International Centre for Economics and Related Disciplines.

Blundell, R., Bond, S., Devereux, M. and Schiantarelli, F. (1992) Investment and Tobin's q: evidence from company panel data, *Journal of Econometrics* **51**: 233-257.

Blundell, R. and Bond, S. (1998) Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* **87**: 115-143.

Breitung, J. (2000) The local power of some unit root tests for panel data, in B. Baltagi (ed.), *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Amsterdam: JAI Press: 161-178.

Breitung, J. and Pesaran, H. (2005) Unit Roots and cointegration in panels, http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0535.pdf (November 4, 2005).

Cable, J.R. (1997) Market share behavior and mobility: an analysis and time-series application, *The Review of Economics and Statistics* **79**: 136-141.

Caselli, F. , Esquivel, G. and Lefort, F. (1996) Reopening the convergence debate: a new look at cross-country growth empirics, *Journal of Economic Growth* **3**: 363-389.

Choi, I. (2001) Unit root tests for panel data, *Journal of International Money and Finance* **20**: 249-272.

Das, B.J., Chappell, W.F. and Shughartii, W.F. (1993) Advertising, competition and market share instability, *Applied Economics* **25**: 1409-1412.

Davidson, R. and MacKinnon, J.G. (1993) *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Dickey, D. A. and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**: 427-431.

Durlauf, S., Johnson, P. and Temple, J. (2004) Growth econometrics, Social Systems Research Institute, SSRI *Working Papers 18*. http://www.ssc.wisc.edu/econ/archive/wp2004-18.pdf (February 6, 2006).

Elliott, G., Rothenberg, T. and Stock, J.H. (1996) Efficient tests for an autoregressive unit root, *Econometrica* **64**: 813-836.

European Commission (2003), Intra- and extra-EU trade, Annual data, Combined Nomenclature, Supplement 2, EUROSTAT, CD ROM of COMEXT trade database.

EU Commission (2005) The EU relations with Chile, http://europa.eu.int/comm/external_relations/chile/intro/ (November 24, 2005).

EViews 5: User's Guide (2004), Quantitative Micro Software, LLC, Irvine, CA.

FAO Production Yearbook (2003) Food and Agricultural Organization of the United Nations, Rome, Vol. 57.

FAO Trade Yearbook (2003) Food and Agricultural Organization of the United Nations, Rome, Vol. 57.

FERN, Greenpeace, WWF 2004:
http://www.panda.org/news_facts/newsroom/news.cfm?uNewsID=17214 (July 28, 2005).

Granger, C. and Newbold, P. (1974) Spurious regressions in econometrics, *Journal of Econometrics* **2**: 111-120.

Greene, W.H. (2000) *Econometric Analysis*, London: Prentice Hall International (UK) Limited.

Hadri, K. (1999) Testing the null hypothesis of stationarity against the alternative of a unit root in panel data with serially correlated errors, *The Econometrics Journal* **3**: 148-161.

Hadri, K. (2000) Testing for stationarity in heterogeneous panel data, *Econometric Journal* **3**: 148-161.

Hadri, K. and Larsson, R. (2005) Testing for stationarity in heterogeneous panel data where the time dimension is finite, *The Econometrics Journal* **8**: 55-69.

Hamilton, J.D. (1994) *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Hayashi, F. (2000) *Econometrics*, Princeton, NJ: Princeton University Press.

Holtz-Eakin D., W. Newey and H. Rosen (1988) Estimating vector autoregressions with panel data, *Econometrica* **56**: 1371-1395.

Hujer, R., Rodrigues, P. and Zeiss, C. (2005) Serial correlation in dynamic panel data models with weakly exogenous regressors and fixed effects, *Working Paper*, March 9, 2005, J.W. Goethe-University, Frankfurt, Germany.

Hula, D.G. (1989) Intangible capital, market share and corporate strategy, *Applied Economics* **21**: 1535-1547.

Im, K., M. Pesaran and Y. Shin (2003) Testing for unit roots in heterogeneous panels, *Journal of Econometrics* **115**: 53-74.

Islam, N. (1995) Growth empirics: a panel data approach, *Quarterly Journal of Economics* **110**, 1127-1170.

2002 Japan Conference: A Summary of the Papers, NBER Website; http://www.nber.org/2002japanconf/sutton.html (July 29, 2005).

Keane, M. and Runkle, D. (1992) On the estimation of panel data models with serial correlation when instruments are not strictly exogenous, *Journal of Business and Economic Statistics* **10**: 1-9.

Kelejian, H.H. and Oates, W.E. (1981) *Introduction to Econometrics. Principles and Applications*, New York: Harper & Row Publishers.

Kim, M.K., Cho, G.D. and Koo, W.W. (2003) Determining bilateral trade patterns using a dynamic gravity equation, Agribusiness & Applied Economic Report No. 525, Center for Agricultural Policy and Trade Studies. North Dakota State University. http://www.ag.ndsu.nodak.edu/capts/documents/NovemberNewsletter2003-2-4-04.pdf (March 3, 2006).

Koyck, L.M. (1954) *Distributed Lags and Investment Analysis*, Amsterdam: North Holland.

Levin, A., Lin, C.F., and Chu, C. (2002) Unit root tests in panel data: asymptotic and finite sample properties, *Journal of Econometrics* **108**: 1-24.

Maddala, G.S. and Wu, S. (1999) A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics* **61**: 631-652.

Ng S. and Perron, P. (2001) Lag length selection and the construction of unit root tests with good size and power, *Econometrica* **69**: 1519-1554.

Phillips, P. and Perron, P. (1988) Testing for a unit root in time series regression, *Biometrika* **75**: 335-346.

OECD (1997) *The Uruguay Round Agreement on Agriculture and Processed Agricultural Products*, OECD Publications, Paris.

Pedroni, P. (1999) Critical values for cointegration tests in heterogeneous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics* **61** Suppl.: 653-670.

Pedroni, P. (2004) Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis, *Econometric Theory* **20**: 597-625.

Rao, B.B. (2007) Estimating short and long-run relationships: a guide for the applied economist, *Applied Economics* **39**: 1613-1625.

Resende, M. and Lima, M.A.M. (2005) Market share instability in Brazilian industry: a dynamic panel data analysis, *Applied Economics* **37**: 713-718.

Schmidt, P., Ahn, S.C. and Wyhowski, D. (1992) Comment, *Journal of Business and Economic Statistics* **10**: 10-14.

Sevestre, P. and Trognon, A. (1996) Dynamic Linear Models, in *The Econometrics of Panel Data. A Handbook of the Theory with Applications,* edited by L. Mátyás and Sevestre, P. 120-144, Dordrecht: Kluwer Academic Publishers, 2nd ed.

Stock, J. (1994) Unit roots, structural breaks, and trends, Chap. 46 in *Handbook of Econometrics*, Vol. IV, edited by R. Engle and D. McFadden, Amsterdam: Elsevier.

Stock, J.H. and M.W. Watson (2003) *Introduction to Econometrics*, Boston […]: Addison Wesley.

Sutton, J. (2004) Market share dynamics and the 'Persistence of Leadership' Debate. The Economics of Industry Group/Suntory and Toyota International Centres for Economics and Related Disciplines, London School of Economics, 37.

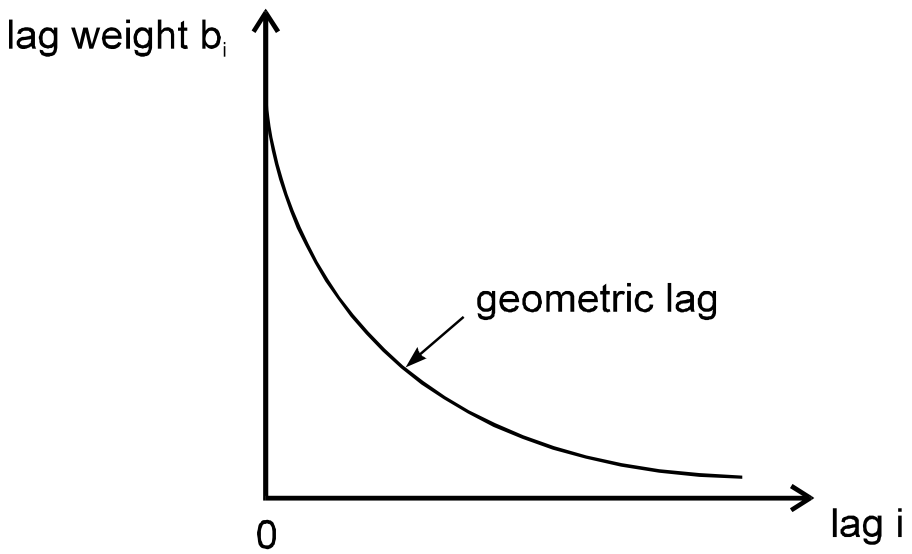
World Bank (2002) TradeCAN (Competitiveness Analysis of Nations) 2002 CD-ROM, Washington, D.C.

World Bank (2005) *World Development Indicators,* Data on CD ROM, Washington, D.C.

WTO Trade Policy Review, European Union (1995, 1997, 2000), World Trade Organisation, Geneva.

WWF (World Wildlife Fund): EU imports of wood-based products 2002 (2005): http://www.panda.org/about_wwf/wherewework/europe/problems/illegal_logging/ (April 29, 2005)

Ziliak, J. (1997) Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators, *Journal of Business and Economic Statistics* **15**: 419-431.



Notes: This reaction pattern must apply to all regressors (x_1, x_2, x_3,…., x_p) and at all levels (cross-sections i)! λ is assumed to be the same for all regressors.

Table 1. *Results for the ARDL market-share model estimated by 3SFGLS (with fixed effects)*

Sector results	Regression coefficients Equation 8				Goodness of fit measure		
	lreer β_0	lreer* γ_0	Adjustm. coeff. (λ)	AR-term	(weighted) R^2adjusted[1]	SE	DW
03 short run	0.82**	-0.72	-0.19	0.68***	0.97	1.02	2.15
	(0.02)	(0.19)	(0.20)	(0.00)			
08 short run	1.82**	-0.14	-0.07	0.69***	0.99	1.05	1.99
	(0.02)	(0.85)	(0.70)	(0.00)			
22 short run	-2.09***	2.01***	0.62***	-0.08	0.98	1.05	2.04
	(0.01)	(0.01)	(0.00)	(0.64)			
22 long run	-5.50***	5.29***					
26 short run	1.83***	0.06	0.70***	-0.29*	0.96	1.02	2.06
	(0.00)	(0.42)	(0.00)	(0.07)			
26 long run	6.10***	0.20					
44 short run	0.35	-2.35	0.46***	0.60***	0.94	1.06	2.36
	(0.76)	(0.13)	(0.00)	(0.00)			
44 long run	0.65	-4.37					
47 short run	-1.20***	-0.27	0.37***	0.01	0.99	1.07	1.87
	(0.00)	(0.42)	(0.00)	(0.91)			
47 long run	-1.90***	-0.43					
74 short run	-0.45***		0.80***	-0.07	0.99	1.04	2.16
	(0.00)		(0.00)	(0.66)			
74 long run	-2.25***						

Notes: p-values in brackets. *, **, *** denote significance at the 10, 5, and 1-percent level, respectively. The estimated values of the fixed effects are not reported. [1]In 3SLS the adjusted R^2 is negative at times. It is unclear how the goodness of fit measures of the different cross-sections are to be weighted in order to derive an overall goodness-of-fit measure. Therefore, the figures listed should only signal the trend.

Table 2. *Results for the ARDL market-share model estimated by non-standard panel GMM (with fixed effects)*

Sector results	Regression coefficients Equation 8				Goodness-of-fit measures		
	lreer β_0	lreer* γ_0	Adjustm. coeff. (λ)	AR-term	(weighted) R^2adjusted	SE	DW
03 short run	-0.20	-0.78***	0.64***	-0.24**	0.98	1.04	2.11
	(0.24)	(0.00)	(0.00)	(0.02)			
03 long run	-0.55	-2.17***					
08 short run	2.29*	-0.15	-0.15	0.69***	0.99	1.10	1.98
	(0.07)	(0.90)	(0.42)	(0.00)			
22 short run	-2.53***	2.29***	0.58***	-0.13	0.98	1.06	2.08
	(0.00)	(0.00)	(0.00)	(0.41)			
22 long run	-6.02***	5.45***					
26 short run	0.12	-0.28***	0.89***	-0.21***	0.87	1.09	2.05
	(0.69)	(0.01)	(0.00)	(0.05)			
26 long run	1.09	-2.54***					
44 short run	-1.22**	-0.98	0.74***	-0.37***	0.82	1.06	2.26
	(0.04)	(0.14)	(0.00)	(0.00)			
44 long run	-4.69**	-3.77					
47 short run	-1.07**	-0.31	0.40***	-0.05	0.74	0.26	1.87
	(0.05)	(0.52)	(0.00)	(0.80)			
47 long run	-1.78**	-0.52					
74 short run	-1.45**	--------	0.37***	0.49***	0.99	1.18	2.01
	(0.02)		(0.03)	(0.00)			
74 long run	-2.30						

Note: p-values in brackets. *, **, *** denote significance at the 10, 5, and 1-percent level, respectively.
[1]The estimated values of the fixed effects are not reported.

Table 3. *Error analysis in the 3SLS and the non-standard GMM framework*

Sector	3SLS framework			Non-standard GMM framework		
	Computed adjustment coefficient λ_{3SLS}	Error if k_{max} =1: λ^2_{3SLS}	Error if k_{max} =2: λ^3_{3SLS}	Computed adjustment coefficient λ_{GMM}	Error if k_{max}=1: λ^2_{GMM}	Error if k_{max}=2: λ^3_{GMM}
Fish (03)	-0.19			0.64***	0.41	0.26
Fruit (08)	-0.07			-0.15		
Beverages (22)	0.62***	0.38	0.24	0.58***	0.34	0.20
Ores (26)	0.70***	0.49	0.34	0.89***	0.79	0.70
Wood (44)	0.46***	0.21	0.10	0.74***	0.55	0.40
Wood pulp (47)	0.37***	0.14	0.05	0.40***	0.16	0.06
Copper (74)	0.80***	0.64	0.51	0.37***	0.14	0.05

Notes: *, **, *** denote significance at the 10, 5 and 1-percent level, respectively. The adjustment coefficients λ_{3SLS} and λ_{GMM} are taken from tables 1 and 2, respectively.