Opportunity costs and gains from economic integration: What happens with the Spanish agricultural sector?

Alicia Gómez Tello

This version: June 2014
(Preliminary and incomplete: please do not circulate)

Abstract

This study aims at assessing the importance of productivity differences as a driving force for land specialization. We refer to a canonical Ricardian setting in which the opportunity costs and the patterns of specialization may change over time. We propose an empirical study for the Spanish agricultural sector by taking into account 50 provinces, 25 crops and the period 1975-2011. After having established the efficient land allocation, we compare the potential and the actual crop outputs. Our preferred regression shows that there exists a positive and statistical significant elasticity between the potential and the actual outputs, even if far from the theoretical predicted value. We investigate this issue and we identify a critical breakpoint for our results at the moment Spain joined the European Economic Community.
1 Introduction

The Ricardian model predicts that regions produce and export goods in which they are relatively more productive. As consequence, the economic integration exacerbates specialization since it drives regions to specialize in those goods in which they are relatively more efficient in production. Then, any device that fuels productivity is also expected to affect competitiveness. The key message of the Ricardian prediction is that productivity differences play an important role in the patterns of specialization.

In this type of setting, this study aims at investigating the importance of productivity differences in the specialization in Spain before and after joining the European Economic Community (EEC) in 1986. Our strategy is to look at land heterogeneity to provide novel predictions about regional specialization. Our theoretical framework of reference is the model developed by Costinot (2009) which considers multiple regions, goods, and factors of production to determine the patterns of international specialization according to the idea of the comparative advantage.\textsuperscript{1}

The Ricardian model is one of the most known and distinguished theories in economics but it is difficult to test empirically. The main problem of testing the Ricardian model is that the relative productivity differences are not always observable because not all the goods are produced in all the regions. A good could not be produced in a specific region because its productivity is very low, its price is very low, or both. The nature of the agricultural sector, however, makes possible through an agronomic model to establish the productivity of all the crops in any part of the world (no matter that one crop is not really produced in a specific area). The pioneer contribution by Costinot and Donaldson (2012) evaluates the Ricardian idea of comparative advantage using a sample of 55 countries and 17 crops in the year 1989. They compute a predicted output by country and crop assuming that each parcel of land is specialized in the crop that maximizes the land revenue. Then, they estimate the relationship between the predicted and the actual production. By regressing the real output against the predicted one (in logarithms) they determine the presence of a positive and significant elasticity, meaning that the Ricardo’s theory of comparative advantage has a significant explanatory power in their analysis. Thus the actual production and the land specialization approximate to those hypothetically predicted by the model. Put differently, this elasticity can be also read as a measure of the efficiency of a current production system: when it is close to one it implies that current production system converges versus the efficient production system which maximizes

\textsuperscript{1}Dorbusch, Fischer, and Samuelson (1977) are the first to extend the standard Ricardian model developing a model with a continuum of goods.
the profitability of the available resources.

Following Costinot and Donaldson (2012)’s methodology, we propose an empirical study for the Spanish agricultural sector by accounting for 50 provinces, 25 crops and the period 1975-2011. One of our challenges is to elaborate an empirical strategy to associate the productivity of the selected crops to each piece of land. The decision to select the agricultural sector is twofold. On the one hand, the agriculture is one of the sectors that better fits the assumptions of the Ricardian model because the sources of comparative advantages keep quite unchanged over time. Of course, important changes have been introduced into the production process (new chemicals, new machinery, and better technology) but if we consider a medium-term period (40 years) these changes has been less radical than in other sectors. On the other hand, this sector is particular relevant when thinking of the effects of the possibility to benefit from the policies implemented by the European Union (EU) once joined the union. In this respect, the common agricultural policy (CAP) represent around 40 percent of the annual EU budget. One of the most important purposes of the CAP is to improve the productivity and the competitiveness of the sector. Some studies have been produced with the purpose to assess the impact of the CAP in the Spanish agriculture by considering a specific change of the policy on a specific crop or region (for instance, Paniagua Mazorra (2001), Oñate et al. (2007)). Instead in this paper, we attempt to assess in what extent the CAP has deviated the Spanish specialization in production from the efficient scenario.

In order to be able to get our objective, namely to assess in which extent the productivity differences are driven the regional specialization in the agricultural production, we need to define the potential output. This potential output is obtained from an hypothetical scenario where the available resources are employed in those products that maximize revenues. Specifically, in this paper we consider the land as an heterogeneous factor of production. The total surface is divided in different parcels of land. Crops can be produced in all the parcels but the productivity depends on the type of land as well as the climate of the area where is situated the parcel. In our hypothetical scenario each parcel of land is used to produce the crop that brings the highest land revenue which is equal to productivity times price. As we assume an unique market price for each crop (this means that the prices are the same for all the provinces), the land heterogeneity across provinces will be the key for regional specialization in the agricultural production.

Once the efficient factor allocation has been determined, namely what to produce in each parcel of land, the potential output by crop, province, and year can be calculated. Then, in order to assess how efficient is the current production system in the agricultural sector, we need
to compare the potential and the actual outputs. A priory we can identify one reason that can explain the deviation between these two variables. While the potential output is obtained from an hypothetical scenario where the land allocation is the efficient one, the actual output could be influenced by specific policies and external circumstances that are not captured in the theoretical model and affect the production decisions.

This empirical exercise relies on an original database. This database has been built using two main resources: the data for productivity have been extracted from the Global Agro-Economic Zones (GAEZ) project while data about prices and actual production are extracted from the Anuario de Estadística Agraria elaborated by the Ministerio de Agricultura, Alimentación y Medio Ambiente. Using and agronomic model and information concerning climate, soil, land cover, and protected areas, the GAEZ project reports information about the potential productivity (kg/ha) of each crop. One of the most important characteristics of the GAEZ database is its level of precision. The Spanish territory is divided into 9,977 grid-cells and each grid-cell is considered as a distinct parcel of land. For each crop-parcel pair a specific level of productivity is provided. Furthermore, GAEZ reports different values of productivity depending on the assumptions chosen concerning water supply and level of technology exploited in the agricultural production. Thus, we can approximate the hypothetical scenario as much as possible the real one taking into account several factor that might determine the actual productivity.

According to our preliminary results, our preferred regression shows that in the case of the non-irrigated crops there exists a positive and statistically significant elasticity between the potential and the actual outputs. This elasticity is equal to 0.08, so it is very far from the potential value calculated from the maximization of the available resources. Furthermore, we identify a critical breakpoint for our results at the moment Spain joined the EEC. Since 1986 the actual output moved away from the predicted efficient scenario. In the case of the irrigated crops we have not found any significative relationship between the potential and the actual outputs, probably because we are nor approximating well the exploited technology.

The rest of the study is organized as follows. Section 2 reviews the institutional background. Section 3 develops the theoretical model and defines the potential output, namely the efficient scenario that will be compared with the actual one. Section 4 presents the empirical exercise and the database. Section 5 shows the preliminary results. Finally, Section 6 concludes and proposes some further extensions.
2 Institutional Background

In order to provide a reasonable interpretation of the empirical results, it is important to describe briefly some distinguishing features of the Spanish agricultural sector and the corresponding communitarian policy.

According to Barciela et al. (2005) the crisis of the traditional agriculture in Spain began at the beginning of the 1970s. Chemicals, fertilizers, and machinery were introduced in the production process. The human and animal strength were replaced by tractors, combines harvested, and other mechanical means of production. There was also important changes in the land distribution. On the one hand, the number of parcels decreased but its median size increased. On the other hand, there was an increase in the use of irrigation techniques due to the higher availability of water resources and a change in the demand toward irrigated crops.

Spain joined the EEC in 1986. The Spanish membership was expected to induce Spanish activities to face the challenges of a large and integrated market, also for the agricultural products. Spain was characterized by its production of fruits (specially citrus), vegetables, and wine. Nevertheless, the trade of the agricultural products during the second half of the 1980s was not balanced: Spanish imports from the EEC-9 increased by 245 percent while Spanish exports to the EEC-9 increased only by 57 percent (Fearne (1997), p. 49). As a reason, the EEC-9 members implemented strict controls from the Spanish exports in side the EEC-9, and this could be one of the reasons of this asymmetry.\footnote{A special monitoring system was introduced to monitor and control the trade in those products identified as ‘sensitive.’ It was known as the \textit{Supplementary Trade Mechanism} (Fearne (1997)).}

Being a member of the EEC means to enjoy the common market and the communitarian policies, among other the CAP. The common market created in the Treaty of Rome (1957) also included the agricultural products. The article 39 specified five objectives for the common agriculture policy: to increase productivity, to ensure a fair standard of living for the agricultural population, to stabilize markets, to guarantee regular supplies, and to ensure reasonable prices in supplies to consumers. While the objectives were detailed, the specific mechanism by which were to be achieved were not specified.

Officially the CAP was established in 1962. It started as a price support policy with the objective to maintain agricultural prices high and stable. Each year the Council of Ministers set a \textit{target} price for each product as a reference point for farmers and politicians. The price support policies consisted in keeping internal prices near to the \textit{target} prices. For instance, import \textit{levies} (i.e., import taxes) were imposed in order to ensure that imports never pushed the internal EEC
prices down below a defined price floor. Different Common Market Organizations (CMOs) were created as a way to manage the regulated products (namely, cereals, sugar, dairy, beef, wine, and olive oil). Each Common Market Organization (CMO) defined a set of rules concerning quality requirements and market price support. Other products (eggs, poultry meat, pork, fruits, and vegetables) also have been protected by external markets but the price interventions were more limited. Products as the seed potatoes were considered free products.

One of the most drastic consequence of the price support policy was that the production of supported products grew faster than the consumption. Thus, the EEC started to buy agricultural products and part of these surpluses were dumped on the world market. Two important problems appeared during the decades of the 1970s and 1980s: the budgetary problems of the agricultural products and the pressure of third countries (above all net exporters countries) due to the effects of the communitarian policy on the agricultural world market. One alternative to solve theses problems and to keep the agricultural incomes high was to implement direct income payments instead of price support policies.

With the Mac Sharry reform (1992) and the Agenda 2000 important changes were introduced: price reductions, farm income compensations, volume restrictions, and market oriented measures, among others. Although the direct payments (per hectare or animal) increased steadily, they were not decouple from current levels of farm production, generating still important distortions in the production. The decoupling of direct payments started with the Fischler reform (2003). By 2006, 82 percent of EU direct payments were decoupled from specific output or resource levels (Jongeneel and Brand (2010), p. 193). Furthermore, the Council of Ministers established the Single Payment Scheme (SPS) to manage the direct payments for farmers and landowners. These direct payments were conditional on some minimal standard requirements concerning food safety, hygiene, environment, animal welfare, and land management. The aim was, on the one hand, to contribute to the development of sustainable agriculture and, on the other hand, to make the CAP more compatible with the expectations of the society (Jongeneel and Brand (2010)).

---

3 The price floor is the minimum import price and it was set below the target price. The price floor was equal to the world price plus the import levy. As the prices of the agricultural products are instable, the variable levy should frequently adjust in order to keep the price floor stable. For a further discussion see Baldwin and Wyplosz (2012), Chapter 9 and Ritson (1997).

4 These reforms were essential to complete the Uruguay Round (1986-1994) and, then, to reduce the international pressure.

5 The EU15 countries introduced this system between 2005-2007. A maximum amount destinated to direct payments was established for each country. There were three options to decide how to distribute the direct payments. Spain applied the historic approach, i.e., the amount received for each farmer’s entitlement corresponded to the payment received during a reference period and the number of hectares then used.
Since 2007 until now important reforms have been introduced. In 2007 the Council of the EU passes the creation of a new single CMO (Regulation 1234/2007).\textsuperscript{6} This single CMO was fully active since 2009 and it include the following products: cereals, rice, sugar, dried fodder, seeds, hops, olive oil and table olives, flax and hemp, bananas, live plants and flowers, raw tobacco, beef and veal, milk and dairy products, pigmeat, sheep and goat meat, eggs, poultry meat, fruits and vegetables and wine (Silvis and Lapperre (2010), pp. 179-180).

Currently the objectives of the CAP can be divided into two pillars. The first pillar concerns about the income support for the farmers (the implemented policies are: market interventions, coupled subsidies, and direct income support) while the second pillar covers matters related to the rural development, the environment, forest, and fisheries. The second pillar has gained relative importance over time\textsuperscript{7} and, consequently, the policies affecting market prices and production decisions have became less important.

3 Theoretical framework

Our theoretical framework grants to Costinot (2009) and Costinot and Donaldson (2012). The aim of this model is to determine the most efficient way to allocate the factor of production, the land. We consider an economy with \( r = 1, \ldots, R \) regions, \( c = 1, \ldots, C \) crops, and \( f = 1, \ldots, F \) heterogeneous factor of production, where each \( f \) corresponds to a distinct parcel of land. A parcel of land \( f \) can be used to produce any type of crop but its return (its productivity) is different. We define \( A_{rf}^c \geq 0 \) as the productivity (kg/ha) of crop \( c \) in region \( r \) and parcel \( f \).\textsuperscript{8} Each region is endowed with a fixed number of parcels. \( L_{rf} \geq 0 \) denotes the endowment (measured in hectares) of parcel \( f \) in region \( r \). The target of the model is to define the best factor allocation across different crops by exploiting all the available land in each region,\textsuperscript{9} by focusing on the supply-side of the economy only. For this reason, it considers a competitive market composed by several farmers that take selling prices \( (p^c \geq 0) \) as given and the technology exhibits constant return to scale.\textsuperscript{10} The total output of crop \( c \) in region \( r \) is given by

\[
Q_r^c = \sum_{f=1}^{F} A_{rf}^c L_{rf}^c ,
\]

\textsuperscript{6}Until 2007 the EU operated with 21 separate CMOs.
\textsuperscript{7}In 2011 the first pillar represents around 30 percent of the total budget while the second pillar represents around 10 percent.
\textsuperscript{8}\( A_{r1}^1 > A_{r2}^1 \) means that the parcel \( f_1 \) is more productive in producing crop 1 (\( c_1 \)) than crop 2 (\( c_2 \)).
\textsuperscript{9}Land allocation is defined as ‘which crop to produce in each parcel.’
\textsuperscript{10}See Appendix A for a further discussion of the assumptions of the model.
where \( A_{rf}^c \geq 0 \) is the productivity (kg/ha) of crop \( c \) in region \( r \) and parcel \( f \) and \( L_{rf}^c \) are the hectares of parcel \( f \) allocated to crop \( c \) in region \( r \). \( L_{rf}^c \) is an endogenous variable, it represents the land allocation across crops. We assume full specialization: given a region \( r \) a parcel of land \( f \) is specialized in only one crop.

The efficient land allocation can be determined solving the following maximization problem

\[
\max_{L_{rf}^c} \sum_{r=1}^{R} \sum_{c=1}^{C} p^c Q_{rf}^c \\
\text{st } \sum_{c=1}^{C} L_{rf}^c \leq \mathcal{L}_{rf}.
\]

Then, the efficient land allocation should fulfill the following two requirements: (1) it maximizes the value of production in all regions, and (2) it has to be feasible.\(^{11}\) The assumptions of the model makes its solution relatively easy: each parcel of land \( f \) should be employed in the crop that maximizes land revenue \((p^c A_{rf}^c)\) independently from the crops that are cultivated in other parcels.

Assuming that the efficient factor allocation is unique,\(^{12}\) the potential (or predicted) output of crop \( c \) in region \( r \) can be expressed as

\[
[Q_{rf}^c]^* = \sum_{f \in \Omega_r^c} A_{rf}^c L_{rf}^c,
\]

where \( \Omega_r^c \) is the set of parcels where the crop \( c \) is produced and it is defined as follows

\[
\Omega_r^c = \left\{ f = 1, \ldots, F \mid \frac{A_{rf}^c}{A_{rf}^c} > \frac{p^c}{p^f} \text{ if } c \neq c' \right\}.
\]

In other words, the crop \( c \) is produced only in those parcels in which it brings the maximum land revenue.

Equations (3) and (4) capture the idea that land heterogeneity plays an important role in determining the patterns of regional specialization. Thus, the type of land determines the specialization in production in each regions. As Costinot and Donaldson (2012) conclude ‘relative rather than absolute productivity differences determine factor allocation and spatial specialization.’ In order to assess empirically the theoretical predictions of regional specialization in

\(^{11}\) We can maximize total profits instead of total revenues. Nevertheless, it does not affect the solution of the problem because the price of a parcel, \( rent_{rf} \), is exogenously given, this means that it is independent of which crop is being produced in parcel \( f \). Furthermore, we assume that \( p^c \), the price perceived by the farmers for crop \( c \), is a net price.

\(^{12}\) This means that for each parcel of land \( f \), there is only one crop that maximizes total revenues. Costinot (2009) reports that the uniqueness is more likely in economies with a large number of regions or a large number of factors. Specifically, in our numerical exercise the efficient land allocation is unique.
production, we compare the potential and the actual outputs in the Spanish agricultural sector for the period 1975-2011.

Our study is simple to hold in an open economy. There is not any constraint that imposes that in a given region \( r \) all the crops (the \( C \) crops) should be produced. Therefore, it is explicitly assumed that if a region is not producing a specific crop \( c \), it could be imported from other regions.

4 Identification strategy

In order to plan our identification strategy we need to have in mind that the key of the Ricardian model is the relative productivity differences. Nevertheless, the main problem of testing the Ricardian prediction is that the relative productivity differences are not always observable because not all the goods are produced in all the region. A good could not be produced in a specific region because its productivity is very low, its price is very low, or both. The nature of the agricultural sector, however, makes possible through an agronomic model to establish the productivity of all the crops in any part of the world (no matter that one crop is not really produced in a specific area).

In order to be able to get our objective, namely to assess in which extent the land heterogeneity is driven the regional specialization in agriculture, we need to determine the potential output. This potential output is calculated by assuming that the factor of production (namely, the land) is used in the most efficient way: its means maximizing land revenue. We should think that the land is heterogeneous; for any crop the productivity depends on the characteristics of the land (nutrients, slope,...) as well as the climate conditions of the area (precipitation, frequency of wet days, mean temperature, temperature range,...). In order to determine these values, the total surface is divided in different parcels of land and for each parcel-crop pair a level of productivity is established. Each parcel of land will specialize in the crop that generates the maximum land revenue. To calculate the revenue we need to know the productivity (kg/ha) but also the market price (Euro/kg). This price should be net, meaning excluding subsidies and taxes. We assume that the price is unique, it means that the price is the same in all the provinces. In this sense, the land heterogeneity across provinces will be the key determinant for regional specialization.

Once the pattern of production have been defined, namely what to produce in each parcel of land, the potential output by crop, province, and year can be calculated. Then we need to compare the potential and the actual outputs. A priory we can identify a few reasons that could
have affected the deviation between these two variables over time. While the potential output is obtained from a theoretical model, the actual output could be influenced by specific policies and external circumstances that are not captured in the theoretical model presented in the previous section.

In order to implement this empirical exercise we build a novel database for the Spanish agricultural sector by province, considering 25 crops and the period 1975-2011.\textsuperscript{13} We choose the year 1975 as the starting point of the analysis because the period 1962-1972 has been characterized by an important transformation: the crisis of the traditional agriculture and the rapid expansion of the industrialization in the Spanish agriculture (Barciela et al. (2005), page 256). We build our database considering two main sources: the Global Agro-Economic Zones (GAEZ) database\textsuperscript{14} and the Anuario de Estadística Agraria (AEA).\textsuperscript{15}

The GAEZ project was launched by the Food and Agriculture Organization (FAO) and the International Institute for Applied Systems Analysis (IIASA) in 2002. The aim of this project is to provide data and knowledge in order to help farmers, researches, and politicians in the implementation of more efficient and sustainable policies concerning the agricultural production. The last and most complete version was published in May 2012 (IIASA/FAO (2012a)).

Relying on an agronomic model and exploiting detailed data about climate, soil, land cover, and protected areas, the GAEZ project delivers information about the potential productivity (kg/ha) of 49 crops in all location around the world. One of the most important characteristics of the GAEZ database is its level of precision. GAEZ divides the Earth in more than 9 million grid-cells (4,320 columns and 2,160 rows). The dimension of the grid is 0.083333 degrees (5 arc-minute), this is approximately 9.28 Km at the equator (as we move away from the equator the dimension of the grid cell decreases). The GAEZ database reports information at grid-cell level.

As an example, Spain is divided in 9,977 grid-cells. The median province contains 192 grid-cells, the largest province (Badajoz) contains 392 grid-cells, and the smallest province (Guipúzcoa) contains only 51 grid-cells (See Appendix B - Table B.1).

To calculate the potential productivity some complex steps and algorithms are implemented. First, the best conditions in terms of climate and land resources are established for each crop. Then, taking into account the agro-climatic and agro-edaphic constrains, the expected potential productivity (kg/ha) of 49 crops in all location around the world. One of the most important characteristics of the GAEZ database is its level of precision. GAEZ divides the Earth in more than 9 million grid-cells (4,320 columns and 2,160 rows). The dimension of the grid is 0.083333 degrees (5 arc-minute), this is approximately 9.28 Km at the equator (as we move away from the equator the dimension of the grid cell decreases). The GAEZ database reports information at grid-cell level.

\textsuperscript{13} These crops are: alfalfa, barley, bean, cabagge, carrot, chickpea, citrus, cotton, grape (to transform), maize, oat, table olives, olive oil, onion, pea (dry), potato (white), rice (wetland), rye, sorghum, soybean, sugar-beet, sunflower, tobacco, tomato, wheat.

\textsuperscript{14} http://gaez.fao.org/Main.html#

\textsuperscript{15} http://www.magrama.gob.es/es/estadistica/temas/publicaciones/anuario-de-estadistica/default.aspx#para3
productivity is calculated by crop and grid-cell. This value is computed using different scenarios concerning water supply (rain-fed or irrigation), the input and technology levels (low, intermediate, or high) and the time period (historic levels, average levels, or future time periods).

Land is the factor of production in our model. We consider each grid-cell as a different parcel of land \( f \) and its total area as the endowment \( L_{rf} \) \( \geq 0 \). The parcels of land have different characteristics and are located in different climate regions, so for a given crop \( c \) the productivity level depends on the parcel of land. It is important to remind that the same parcel of land \( f \) could be in more than one province. The productivity of this parcel will be the same, independently of the province, but the potential output will depend on the area of this specific parcel available in each province.

Our proxy for productivity \( (A^c_{rf}) \) corresponds to two different variables of the GAEZ project: total production capacity \( (TPC) \) and potential production capacity for current cultivated land \( (PPC) \). For the preliminary results presented in this paper we choose the intermediate input level. Concerning the time period we choose the baseline period, it is defined as the average climate conditions for the period 1961-1990. The \( TPC \) reports the potential production capacity in terms of output density (kg/ha) by grid-cell. So, for a given grid-cell different productivity values are calculated, one for each crop. The productivity of a specific grid-cell could be zero for all the crops, meaning that the characteristics of this piece of land are not good for agriculture.

The variable productivity is used to compute the potential output. We try to approximate the potential output to the reality by taking into account several factors that might affect the actual productivity. In this sense, we differentiate from Costinot and Donaldson (2012) distinguishing between irrigated and non-irrigated techniques. Irrigation techniques can affect the level of productivity but the irrigation infrastructure is unevenly distributed across Spain and this is an important variable to take into consideration. The variable \( PPC \) reports the potential average yield (kg/ha) of current cultivated land share of grid-cell and allow us to choose between rainfed or irrigation options. Furthermore, we think that this variables is more realistic because it

---

16For more detailed description see IIASA/FAO (2012b), Module V.

17The data from the GAEZ project is used spread in economic literature. Nunn and Qian (2011) use information about the regional suitability for growing potatoes in countries of the Old World (Africa, Asia, and Europe), Costinot and Donaldson (2012) extract information about the potential production capacity of 17 crops in 55 countries, and Bustos et al. (2013) use data on potential yield for soy and maize in Brazil considering different levels of technology.

18The median grid-cell measures 6,358.61 ha (63.59 square km), the largest grid-cell measures 7,932.99 ha (79.33 square km), and the smallest grid-cell measures only 0.001 ha (0.00001 square km).

19It assumes a partly market oriented system (that is, production for subsistence and commercial sale). Production is medium labor intensive, it means that fertilizers, chemicals, disease and weed control, adequate fallows and some conservation measures are used in the agricultural production.
considers the part of the grid-cell that is currently cultivated.\footnote{The \emph{PPC} reports a value equal to -1 when any share of the parcel of land is being used to the agriculture. Concerning irrigation at least 1 percent of the total parcel of land should have an adequate infrastructure for irrigation.}

In order to determine the efficient land allocation described in equation (4) information about prices is needed. Prices are extracted from our second data source, the AEA. The AEA is published by the \textit{Ministerio de Agricultura, Alimentación y Medio Ambiente} (MAGRAMA) since 1904. The information about prices is at national level and is defined as the ‘price received by the farmer.’\footnote{From 1975 to 1993 the prices are expressed in ptas/kg while from 1994 to 2011 the prices are expressed in euro/100kg. After some simple operation we get euros/kg for all the period (1 euro = 166.386 pesetas).} With information about productivity, endowments, and prices and implementing equation (3) and (4), efficient land allocation and potential outputs can be determined (\textit{Appendix B} explains in detail how to solve the numerical exercise).

Actual or real production is also extracted from the AEA. The information is reported in kilograms or tones, by crop, province, year, dry, and irrigation.\footnote{We have missing values for the following crops and years: rice (1982-1984), maize (1990), cotton (1975-1992, 1995), and sunflower (1975-1995). These crops and these years should be excluded from the analysis, also from the maximization problem.} In order to build correctly this database some corrections and approximation have been needed (See \textit{Appendix C} for the details).

## 5 Preliminary results

According to the identification strategy previously explained, first we proceed by determining the efficient land allocation. The idea is to maximize the value of the land, particularly each parcel of land specializes in the crop that maximizes revenues (Euro/ha). Then, taking into account the efficient land allocation and the land endowment, a potential output by region \( r \) and crop \( c \) is calculated. An interesting exercise is to compare these predictions with the real data. A perfect correlation between the potential and the actual outputs means a high significance of the Ricardian prediction, namely each region specializes in those crops in which the region is more productive. Nevertheless, the actual output can be influenced by specific policies and other conditions that are not captured in the theoretical model. On the one hand, our objective is to assess the extent Ricardian prediction fits the actual production of the Spanish agricultural sector and, on the other hand, to evaluate the effects of the communitarian policies on the production decisions.

We present three different alternatives that differ according to the information extracted from the GAEZ database. Hence, the numerical exercise described in the \textit{Appendix B} needs
be solved three times, one for each alternative. The Alternative 1 considers the variable total production capacity as a measure of productivity and dry as water supply option. The Alternative 2 considers the variable potential production capacity as a measure of productivity and dry as a water supply option. Finally, the Alternative 3, considers the variable potential production capacity as a measure of productivity and irrigation as water supply. Concerning the level of inputs, the three alternatives consider the intermediate level. It is important to differentiate between non-irrigated and irrigated techniques. First, the productivity (kg/ha) increases with irrigated techniques and, second, the technology used with irrigation is expected to be more sophisticated (in fact, we have in mind to do a new alternative combining irrigation and high inputs). We start discussing some preliminary evidence. Figure 1 shows the differences between dry versus irrigation production in Spain. For each crop the actual production of all the period is normalize to 100. The bars indicate which part of this production is due to dry production and which part is due to irrigation production. We can appreciate that the production of some crops (citrus, rice, and vegetables) comes mainly from the irrigation production system while the production of other crops (rye, chickpea, and oat) comes mainly from the dry production system.

Figure 1: Water supply by crops (Spain, 1975-2011)

Once computed the potential outputs (available upon request) we compare them with the actual values. The results of each alternative are presented in different tables, build according to the same structure. We regress the logarithm of the actual output on the logarithm of the predicted output. Logarithms serve to control in part for heterogeneity. Column (1) is the
baseline specification. As in Costinot and Donaldson (2012) it establishes the first correlation between our variables of interest. Column (2) controls by crop, province, and year fixed effects. In columns (3) and (4) the period is split into two in order to assess if the entrance to the EEC had any effect on the production decisions. Standard errors are clustered by province to account for potential within-province (across crops) correlation.

Table 1 presents the results for the Alternative 1. As dependent variable the actual output of dry production is considered. Column (1) reports an elasticity of 0.15, indicating that there is a positive and statistically significant correlation between the actual and the potential outputs. This value, however, decreases to 0.08 after controlling by fixed effects (column (2)). When the sample is split into two periods, the coefficient for the first period is slightly higher than the coefficient for the second period, but this difference is not statistically significant different from zero. Results for the Alternative 2 are reported in Table 2. As before the dependent variable only include dry production. The results are very similar to those previously presented, although the value of the elasticity is slightly lower in the four regressions. There is a positive and statistically significant relation between the actual and the potential outputs in the baseline regression (column (1)). This elasticity decreases after taking into account fixed effects (column (2)). We perceive a difference before and after the entrance to the EEC, but in this case the difference between the coefficients is statistically significant different from zero at 10 percent. Finally, results for the Alternative 3 are presented in Table 3. As dependent variable only the actual production of the irrigated production is considered. In this case, only column (1) reports a positive and statistically significant coefficient different from zero. Further investigation is required to assess whether this results is associated with the level of technology chosen.

These results allow us to confirm that there are differences between the non-irrigated and irrigated production systems. The proposed model explains better the dry than the irrigated production. At the moment we have detected a slight effect after the entrance to the EEC.
Table 1: Actual versus predicted output (Alternative 1)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>ln(actual output_dry)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(potential output)</td>
<td>0.1453*** 0.0802*** 0.1095*** 0.0851***</td>
</tr>
<tr>
<td></td>
<td>(0.0211) (0.0254) (0.0271) (0.0264)</td>
</tr>
<tr>
<td>constant</td>
<td>7.9763*** 2.4712*** 0.2229 2.1448***</td>
</tr>
<tr>
<td></td>
<td>(0.2210) (0.5015) (0.2220) (0.5029)</td>
</tr>
</tbody>
</table>

**FIXED EFFECTS**
- crop: No Yes Yes Yes
- province: No Yes Yes Yes
- year: No Yes Yes Yes

Observations: 44050 44050 12500 31550
R-square: 0.0167 0.5219 0.5214 0.5325

Standard errors clustered by provinces are in parenthesis
* p < 0.10, ** p < 0.05, *** p < 0.01
GAEZ database
Variable: Total production capacity (TPC)
Inputs: intermediate
Water supply: dry
Weather: baseline period (1961-1990)

Table 2: Actual versus predicted output (Alternative 2)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>ln(actual output_dry)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(potential output)</td>
<td>0.1350*** 0.0572** 0.0957*** 0.0545**</td>
</tr>
<tr>
<td></td>
<td>(0.0201) (0.0236) (0.0254) (0.0241)</td>
</tr>
<tr>
<td>constant</td>
<td>7.8030*** -0.8054 0.9412*** 8.7007***</td>
</tr>
<tr>
<td></td>
<td>(0.2209) (0.5097) (0.2501) (0.7542)</td>
</tr>
</tbody>
</table>

**FIXED EFFECTS**
- crop: No Yes Yes Yes
- province: No Yes Yes Yes
- year: No Yes Yes Yes

Observations: 44050 44050 12500 31550
R-square: 0.0167 0.5219 0.5214 0.5325

Standard errors clustered by provinces are in parenthesis
* p < 0.10, ** p < 0.05, *** p < 0.01
GAEZ database
Variable: Potential production capacity (PPC)
Inputs: intermediate
Water supply: dry
Weather: baseline period (1961-1990)

Furthermore, as a novelty, we can exploit the temporal dimension of our sample. One assumption of the Ricardian model is that technology is constant over time, then also the comparative advantages are constant over time. Although the technology, defined as the potential
Table 3: Actual versus predicted output (Alternative 3)

<table>
<thead>
<tr>
<th>Dependent variable: ln(actual output_{irrigation})</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(potential output)</td>
<td>0.0450*</td>
<td>0.0191</td>
<td>0.0203</td>
<td>0.0263</td>
</tr>
<tr>
<td></td>
<td>(0.0235)</td>
<td>(0.0182)</td>
<td>(0.0179)</td>
<td>(0.0232)</td>
</tr>
<tr>
<td>constant</td>
<td>9.4024***</td>
<td>4.1746***</td>
<td>2.9173***</td>
<td>10.2317***</td>
</tr>
<tr>
<td></td>
<td>(0.4823)</td>
<td>(0.7476)</td>
<td>(0.7579)</td>
<td>(0.7347)</td>
</tr>
</tbody>
</table>

**FIXED EFFECTS**

<table>
<thead>
<tr>
<th></th>
<th>crop</th>
<th>province</th>
<th>year</th>
<th>Sample</th>
<th>Observations</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>1975-2011</td>
<td>44050</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>1975-1985</td>
<td>12500</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>1986-2011</td>
<td>31550</td>
</tr>
</tbody>
</table>

* Standard errors clustered by provinces are in parenthesis

* p < 0.10, ** p < 0.05, *** p < 0.01

GAEZ database
Variable: Potential production capacity (PPC)
Inputs: intermediate
Water supply: irrigation
Weather: baseline period (1961-1990)

productivity (kg/ha) of crop c in region r and parcel f (A^c_{rf}), is constant over time, the changes in prices can vary the efficient land allocation and, therefore, the potential output. We propose a second exercise in which a regression is run every year: there is a comparison between the actual and the potential outputs year by year. Then the coefficients and its corresponding 5 percent confidence intervals are plotted in a graph.

The results for the Alternative 1 are presented in the Figure 2. Before the entrance into the EEC, the coefficient is positive and statistically significant different from zero. The maximum value took place in 1983 and it was equal to 0.15. After the entrance into the EEC, the trend of the coefficient is not so clear. Except for three years (1990, 2005, and 2010) the coefficient is positive and statistically different from zero. The results for the Alternative 2 are plotted in Figure 3. Here a more clear story appears. Before the entrance of the EEC the coefficient is positive and statistically different from zero. The elasticity is around 0.1. After the entrance into the EEC we appreciate that the value of the elasticity is decreasing. From 1999 to the end of the period the coefficient is anymore statistically significant different from zero. Remind that we are comparing the potential and the actual outputs year by year and the potential output is calculated considering that the factor of production is used in the most efficient way. These results suggest that since 1986 the regional specialization has moved away from the efficient scenario. Maybe, some communitarian policies have influenced in the production decision of the
Spanish farmers. The results of the Alternative 3 are not presented in the paper because, except for the last three years of the period, the coefficient is not statistically significant different from zero. Again, we suspect that the level of technology chosen could be the cause of these results.

Figure 2: Actual versus predicted output by years (Alternative 1)

![Figure 2](image1)

Figure 3: Actual versus predicted output by years (Alternative 2)

![Figure 3](image2)

This second empirical exercise allows us to get important results. The differences between the non-irrigated and the irrigated crops have been confirmed and we have detected a distorted impact of the common agricultural policy toward the potential production in the case of the dry production.

Although we found a positive and statistically relationship between the \( \ln(\text{actual output}) \)
and the ln(potential output) in the case of the non-irrigated crops, this dependence is far away from the perfect correlation. We are conscious of some of the limitations in the model. First, the model only considers the land productivity differences as the key of regional specialization (the price of other factors of production or the factor intensity differences are not taking into account). Second, the GAEZ project assumes that there are not regional differences concerning the level of technology of the mean dimension of the parcels of land.

6 Conclusions

This study aims at assessing the importance of productivity differences as a driving force for land specialization. We refer to a canonical Ricardian setting in which the opportunity costs and the patterns of specialization may change over time. We propose an empirical study for the Spanish agricultural sector by accounting for 50 provinces, 25 crops and the period 1975-2011. After having established the efficient land allocation we calculate the corresponding potential output by province, crop, and year. Then, we compare the potential and the actual outputs in order to assess how efficient is the current production system.

The results presented in this paper are preliminary. Our preferred regression shows that in the case of the non-irrigated crops there exists a positive and statistical significant elasticity between the potential and the actual outputs. This elasticity is equal to 0.08, so it is very far from the efficient predicted scenario. Furthermore, we identify a critical breakpoint for our results at the moment Spain joined the EEC. Since 1986 the actual output moved away from efficient scenario. In the case of the irrigated crops we have not found any significative relationship between the potential and the actual outputs. We suspect that this result could be consequence of the level of technology chosen.

For the next version of the paper we should improve several things related to the numerical exercise.

- Concerning the irrigation production we have to consider the option of high technology in order to extract productivity data from the GAEZ.

- Taxes and subsidies should be excluded from the prices. In our current sample, the prices of the sugar-beet and the tobacco consider subsidies until the year 2000. It should be corrected. The website http://cies.adelaide.edu.au/ reports information about the distortions of the agriculture incentives for the period 1955-2007. This information could help us to
‘clean’ the prices of these crops. On the other hand, the World Bank contains information about the world prices of the agricultural products. We can use this database in order to perform a robustness check.

- The AEA contains missing values about the actual production for some specific crops and years: rice (1982-1984), maize (1990), cotton (1975-1992, 1995), and sunflower (1975-1995). These crops and these years should be excluded from the analysis, also from the maximization problem. For instance, if the actual production of the crop $c$ in the year $t$ is missing, the crops $c$ should be excluded from the maximization problem in the year $t$. In this version, however, they have been not excluded from the numerical problem.
References


A Appendix

The objective of this appendix is to provide a quantitative argument about the independence of the efficient factor allocation and the vector of prices from the demand conditions.\textsuperscript{23}

We consider the supply-side of an economy with $C$ commodities. A production vector is defined as $y = (y_1, \ldots, y_C) \in Y^C$, where $y_c > 0$ represents an output, $y_c < 0$ represents an input, and $y_c = 0$ is neither an output nor an input ($\forall c = 1, \ldots, C$). The production set is the set of all production vectors that constitute feasible plants for the firm and it is represented by $Y \subseteq R^C$ ($y \in Y$ is a possible production vector while $y \notin Y$ is a not possible production vector). Given constant returns to scale technology $Y$, the array generated by a vector $\bar{y} \in Y$ is the set:

\[
\{ y \in Y^C : y = \alpha \bar{y} \text{ for some scalar } \alpha \geq 0 \}.
\]

We consider a special case with a list of finitely many activities (say $M$). Each activity is defined as a specific technique to produce the output $y_c$. Then, the production set is defined as

\[
\{ y \in Y^C : y = \sum_{m=1}^{M} \alpha_m a_m \text{ for some scalar } (\alpha_1, \ldots, \alpha_M) \geq 0 \}
\]

where $a_m \in R^C$ is the elementary activity $m$ and $\alpha_m \in R^C$ is the level of the elementary activity $m$.

**Definition 1** Given a price vector $p \in R^C_+$, a profit maximizing plan exist in $Y$ if and only if $pa_m \leq 0$ for every $m$.

If $pa_m < 0$ the level of the elementary activity will be zero ($\alpha_m = 0$); if $pa_m = 0$ the level of the elementary activity will be positive ($\alpha_m > 0$); and if $pa_m > 0$ the profits will be positive and the level of the elementary activity will large.

For any price vector $p >> 0$ generating zero profits, let $A(p)$ denote the set of activities that generate zero profits: $A(p) = \{ a_m : pa_m = 0 \}$. The profit-maximizing supply set $y(p)$ is therefore the convex cone generated by the activities in $A(p)$. The convex profit-maximizing supply set ensures that every efficient $y \in Y$ is a profit maximizing production for some $p >> 0$ (Second Welfare Theorem). A production vector $y \in Y$ is efficient if there does not exist a $y' \in Y$ such that $y' \geq y$ and $y' \neq y$.

We present the **Leontief model** because the efficient input allocation is independent from the demand conditions and the total outputs. The Leontief model considers an economy with $C - 1$ commodities, the commodity $Cth$ is the one primary factor (usually labor), $M$ elementary activities (an elementary activities is a combination of inputs to produce a good, so each activity is a different technique), the technology is constant return to scale, and there is no joint production.

\textsuperscript{23}For a more complete discussion see Mas-Colell et al. (1985), Chapter 5.
In the Leontief model without substitution possibilities a good should be produced using a fixed combination of inputs, then there is one specific elementary activity for each good. In the Leontief model with substitution possibilities, however, a good may be produced with several techniques (then, $a_1 \in R^C, ..., a_{M_x} \in R^C$ is the list of $M_x$ elementary activities capable of producing good $x$). Although substitution of techniques is possible it will never occur because for each good there is associated a single optimal technique. Because of the constant returns to scale technology the optimal technique is independent from the demand conditions and the total level of output. This is known as the non-substitution theorem (Samuelson, 1951). The theorem depends critically on the presence of only one factor of production. With more than one factor of production, the optimal choice of technique should depend on the relative prices of these factors. The cost of the remaining inputs must be calculated in terms of the amount of labor required to produced them. For a given wage rate the relative inputs prices remains unaltered (Chakravarty (2005), page 578). Furthermore, taking into account the Second Welfare Theorem, for a given strictly positive price vector $p >> 0$ an efficient production vector is also a profit-maximizing production. Then, the vector of prices is independent of the preferences and demand conditions.

### Appendix

This appendix provides a complete description on how to determine the efficient land allocation and the potential output.

(a) Let a region $r$ and a parcel of land $f$, the land revenue (Euro/ha) for each crop $c$ is calculated:

<table>
<thead>
<tr>
<th>Crop</th>
<th>Revenues (Euros/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p^{c_1} A_{c_1 f}$</td>
</tr>
<tr>
<td>2</td>
<td>$p^{c_2} A_{c_2 f}$</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>$p^{c_{25}} A_{c_{25} f}$</td>
</tr>
</tbody>
</table>

→ The crop that maximizes the field revenue (Euro/ha) is chose $c^*_j$ ($j = 1, ..., 25$).

→ It is important to check the uniqueness, that is, only one crop maximizes the land revenue.

(b) Given a region $r$, part (a) is repeated for all the parcels of land.

→ For instance, in the case of Valencia ($r = 46$) there are 212 parcels identified as $f_{8984}, f_{8985}, ..., f_{9195}$.
(c) The efficient land allocation is defined: each crop $c$ is produced in a set of parcels of land $(\Omega^c_r)$.

(d) The potential output of crop $c$ in region $r$ is calculated by taking into account the efficient land allocation $\Omega^c_r$ and the endowment $L_{rf}$ (the total surface (ha) of each parcel of land $f$ in region $r$): $[Q^*_c]^t = \sum_{f \in \Omega^c_r} A^c_{tf} L_{rf}$

(e) Parts (a)-(d) are repeated for each year $t$ ($t = 1975, \ldots, 2011$). The efficient land allocation can change because prices are time variant.

Figure B.1: Total production capacity (Kg/ha) - Crop: olive

Source: GAEZ project (version 3.0)
Table B.1: Provinces and grid-cells

<table>
<thead>
<tr>
<th>Province</th>
<th>Autonomous Community</th>
<th>Surface (ha)</th>
<th>id_grid-cells</th>
<th>Total grid-cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alava</td>
<td>País Vasco</td>
<td>303,459.8369</td>
<td>[1, 84]</td>
<td>84</td>
</tr>
<tr>
<td>Albacete</td>
<td>Castilla La Mancha</td>
<td>1,741,742.9791</td>
<td>1,85, 362</td>
<td>278</td>
</tr>
<tr>
<td>Alcalá</td>
<td>Comunidad Valenciana</td>
<td>582,045.6574</td>
<td>365, 358</td>
<td>121</td>
</tr>
<tr>
<td>Almería</td>
<td>Andalucía</td>
<td>876,779.1306</td>
<td>484, 645</td>
<td>165</td>
</tr>
<tr>
<td>Ávila</td>
<td>Castilla León</td>
<td>804,876.7585</td>
<td>649, 811</td>
<td>163</td>
</tr>
<tr>
<td>Badajoz</td>
<td>Extremadura</td>
<td>2,179,003.7990</td>
<td>812, 1207</td>
<td>396</td>
</tr>
<tr>
<td>Islas Baleares</td>
<td>Islas Baleares</td>
<td>301,786.7582</td>
<td>1208, 1335</td>
<td>128</td>
</tr>
<tr>
<td>Barcelona</td>
<td>Cataluña</td>
<td>775,786.2519</td>
<td>1536, 1393</td>
<td>156</td>
</tr>
<tr>
<td>Burgos</td>
<td>Castilla Leon</td>
<td>1,428,151.3400</td>
<td>1492, 1786</td>
<td>285</td>
</tr>
<tr>
<td>Cáceres</td>
<td>Extremadura</td>
<td>1,988,929.4099</td>
<td>1787, 2152</td>
<td>366</td>
</tr>
<tr>
<td>Cadiz</td>
<td>Andalucía</td>
<td>743,995.2857</td>
<td>2133, 2295</td>
<td>143</td>
</tr>
<tr>
<td>Castellón</td>
<td>Comunidad Valenciana</td>
<td>958,963.7643</td>
<td>2386, 2431</td>
<td>136</td>
</tr>
<tr>
<td>Ciudad Real</td>
<td>Castilla León</td>
<td>1,988,929.4099</td>
<td>2432, 2790</td>
<td>359</td>
</tr>
<tr>
<td>Cordoba</td>
<td>Andalucía</td>
<td>1,376,832.0043</td>
<td>2791, 3042</td>
<td>252</td>
</tr>
<tr>
<td>A Coruña</td>
<td>Galicia</td>
<td>769,538.7062</td>
<td>3043, 3212</td>
<td>170</td>
</tr>
<tr>
<td>Lleida</td>
<td>Castilla La Mancha</td>
<td>1,742,017.2519</td>
<td>6213, 5324</td>
<td>312</td>
</tr>
<tr>
<td>Girona</td>
<td>Cataluña</td>
<td>593,705.2901</td>
<td>5255, 3657</td>
<td>133</td>
</tr>
<tr>
<td>Granada</td>
<td>Andalucía</td>
<td>1,265,729.1890</td>
<td>3658, 3857</td>
<td>230</td>
</tr>
<tr>
<td>Guadalajara</td>
<td>Castilla La Mancha</td>
<td>1,220,265.5321</td>
<td>3888, 4125</td>
<td>238</td>
</tr>
<tr>
<td>Guipúzcoa</td>
<td>País Vasco</td>
<td>198,980.9358</td>
<td>4126, 4176</td>
<td>91</td>
</tr>
<tr>
<td>Huelva</td>
<td>Andalucía</td>
<td>1,014,608.4358</td>
<td>4177, 4366</td>
<td>190</td>
</tr>
<tr>
<td>Huesca</td>
<td>Aragón</td>
<td>1,546,720.3124</td>
<td>4367, 4667</td>
<td>301</td>
</tr>
<tr>
<td>Jaén</td>
<td>Andalucía</td>
<td>1,348,663.9149</td>
<td>4668, 4911</td>
<td>244</td>
</tr>
<tr>
<td>León</td>
<td>Castilla Leon</td>
<td>1,559,187.3077</td>
<td>5012, 5020</td>
<td>259</td>
</tr>
<tr>
<td>Lérida</td>
<td>Cataluña</td>
<td>1,219,283.6572</td>
<td>5421, 5451</td>
<td>541</td>
</tr>
<tr>
<td>La Rioja</td>
<td>La Rioja</td>
<td>504,163.0341</td>
<td>5425, 5563</td>
<td>112</td>
</tr>
<tr>
<td>Lugo</td>
<td>Galicia</td>
<td>988,956.3424</td>
<td>5564, 5785</td>
<td>195</td>
</tr>
<tr>
<td>Logroño</td>
<td>Comunidad de Madrid</td>
<td>802,590.2982</td>
<td>5759, 5927</td>
<td>169</td>
</tr>
<tr>
<td>Málaga</td>
<td>Andalucía</td>
<td>730,761.5335</td>
<td>6025, 6076</td>
<td>139</td>
</tr>
<tr>
<td>Murcia</td>
<td>Región de Murcia</td>
<td>1,131,200.4176</td>
<td>6077, 6285</td>
<td>212</td>
</tr>
<tr>
<td>Navarra</td>
<td>Navarra</td>
<td>1,038,558.3242</td>
<td>6259, 6505</td>
<td>217</td>
</tr>
<tr>
<td>Orense</td>
<td>Galicia</td>
<td>720,387.4550</td>
<td>6505, 6695</td>
<td>153</td>
</tr>
<tr>
<td>Asturias</td>
<td>Asturias</td>
<td>1,060,386.7650</td>
<td>6659, 6871</td>
<td>213</td>
</tr>
<tr>
<td>Cantabria</td>
<td>Cantabria</td>
<td>353,583.5671</td>
<td>7477, 7755</td>
<td>77</td>
</tr>
<tr>
<td>Segovia</td>
<td>Castilla León</td>
<td>661,852.8102</td>
<td>7674, 7811</td>
<td>138</td>
</tr>
<tr>
<td>Sevilla</td>
<td>Andalucía</td>
<td>1,404,840.6522</td>
<td>7812, 8099</td>
<td>208</td>
</tr>
<tr>
<td>Soria</td>
<td>Castilla león</td>
<td>1,029,914.1695</td>
<td>8067, 8277</td>
<td>208</td>
</tr>
<tr>
<td>Tarragona</td>
<td>Cataluña</td>
<td>631,486.0099</td>
<td>8278, 8409</td>
<td>132</td>
</tr>
<tr>
<td>Teruel</td>
<td>Aragón</td>
<td>1,481,060.6034</td>
<td>8410, 8691</td>
<td>282</td>
</tr>
<tr>
<td>Toledo</td>
<td>Castilla La Mancha</td>
<td>1,363,119.7445</td>
<td>8954, 9095</td>
<td>222</td>
</tr>
<tr>
<td>Valencia</td>
<td>Comunidad Valenciana</td>
<td>1,081,019.6428</td>
<td>9894, 9195</td>
<td>212</td>
</tr>
<tr>
<td>València</td>
<td>Castilla Leon</td>
<td>810,936.2015</td>
<td>9198, 9306</td>
<td>171</td>
</tr>
<tr>
<td>Vizcaya</td>
<td>País Vasco</td>
<td>221,633.8710</td>
<td>9367, 9427</td>
<td>61</td>
</tr>
<tr>
<td>Zaragoza</td>
<td>Castilla León</td>
<td>1,038,890.3417</td>
<td>9428, 9635</td>
<td>211</td>
</tr>
<tr>
<td>Zaragoza</td>
<td>Aragón</td>
<td>1,727,288.5627</td>
<td>9636, 9977</td>
<td>339</td>
</tr>
</tbody>
</table>

1 Calculus: Author (ArcMap 10)
2 The identification is by grid-cell and provinces
C Appendix

The database exploited in this paper has been built considering two main sources: productivity values are extracted from the GAEZ project elaborated by the FAO and the IIASA while prices and actual outputs are extracted from the AEA published by the MAGRAMA. This appendix details the managing of these two sources in order to get an adequate database to implement the numerical exercise described in the Appendix B.

The first important step is the selection of the crops. The GAEZ database contains information about 49 crops while the AEA database contains information about 150 crops. As we have to merge data from both sources we have to define a the list of crops available in both database and representative for the Spanish agricultural sector. Finally, we work with a sample of 25 crops that are classified in nine groups (Table C.1).

Table C.1: List of crops

<table>
<thead>
<tr>
<th>GROUP</th>
<th>CROPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>fodder</td>
<td>alfalfa</td>
</tr>
<tr>
<td>cereal</td>
<td>barley, maize, oat, (wetland) rice, rye, sorghum, wheat</td>
</tr>
<tr>
<td>citrus</td>
<td></td>
</tr>
<tr>
<td>grape</td>
<td>grape (to transformation)</td>
</tr>
<tr>
<td>legume</td>
<td>(phaseolus) bean, chickpea, (dry) pea</td>
</tr>
<tr>
<td>industrial</td>
<td>cotton soybean, sugar (beet), sunflower, tobacco</td>
</tr>
<tr>
<td>olive</td>
<td>table olives, olive oil</td>
</tr>
<tr>
<td>tubers</td>
<td>(white) potato</td>
</tr>
<tr>
<td>vegetable</td>
<td>cabbage, carrot, onion, tomato</td>
</tr>
</tbody>
</table>

We follow the next challenge: we need to approximate the potential productivity (Subsection C1) and the prices and the actual outputs (Subsection C2) when they are not available.

C.1 Potential productivity

We have to approximate the potential productivity for the grape. The grape is identified by table grapes and grape to transformation (mainly for the elaboration of wine). This group is important in the Spanish agriculture: it represents each year around 6-8 per cent of the total agricultural production.\(^24\) The problem is that the GAEZ database has not information for any of these two varieties. But it is relevant to include at least the grape to transformation.\(^25\)

In order to calculate the potential productivity we choose another type of crop with similar

\(^{24}\text{Source: AEA.}\)

\(^{25}\text{In 1975 the table grapes represented 8.87 percent of the total production of grape while the grape to transformation represented 91.13 percent. In 2011 this differences was even higher: the table grapes represented 4.19 percent of the total production of grape while the grape to transformation represented 95.8 percent (Source: AEA).}\)
characteristics: the olive.\footnote{The olive and the grape follow similar production systems as they are very common crops in the Mediterranean area due to the weather characteristics, temperate winters and hot and dry summers (Espasa (1998)).} First, using information from the AEA we calculate the aggregate actual productivity (kg/ha) for the olive (\textit{prod}^{\text{olive}}) and for the grape to transformation (\textit{prod}^{\text{grape-transf}}).\footnote{\textit{Aggregate} means that we consider all the regions (\(\sum_{n=1}^{50}\)) and all the years (\(\sum_{t=1}^{37}\)).} Second, a ratio between the actual productivity of the olive and the actual productivity of the grape to transformation is calculated in order to establish a relation between both crops, that is, \(r(\text{og}) = \frac{\textit{prod}^{\text{olive}}}{\textit{prod}^{\text{grape-transf}}}.\) Finally, using this ratio and the information reported by the GAEZ about the olive (\(A_{rf}^{\text{olive}}\)), we establish a potential productivity for the grape to transformation, one for each grid-cell \(f\), \(A_{rf}^{\text{grape-transf}} = \frac{A_{rf}^{\text{olive}}}{r(\text{og})}.\) In this way, we can include the grape to transformation in the efficient land allocation problem.

C.2 Prices and actual outputs

Prices are needed in order to determine the efficient land allocation (See Appendix B - part \((a)\)). Unfortunately there are prices that are not available for some specific crops and years. We implement different methods is order to cover the missing values reported by the AEA.

As there are not many differences between the AEA and the FAOSTAT databases, the first method consists in complementing the missing values with data reported by the FAOSTAT.\footnote{http://faostat3.fao.org/faostat-gateway/go/to/download/P/*/E} We use information from the FAOSTAT for the following crops and years: orange (1976-1979), lemon (1976-1979), table olives (1976-1979), potato (1986), and carrot (1975-1984).

Sometimes, the information is neither available in FAOSTAT, then we use the second method: we calculate an average taking into account the last and the first available prices and use this value to cover the years without information. This method is implemented for alfalfa and mandarin. An average price average is calculated using the prices reported for the years 1975 and 1980. Then, this average in used to cover the period 1976-79.

The third method consists of comparing two varieties that belong to the same group (one variety have complete information about prices and the other no). The idea is to use the available prices to calculate a ratio between the prices of the two varieties. Then, this ratio is used as a proportional factor to complete the missing values of the series.

Crops with adjustments in the series are the olive oil and the grape to transformation. The price of the olive oil is available since 1985. In order to complete the period, we use as reference crop the table olives. We focus on the available prices during the 1980s (that is, from 1986-1989). For each year, a ratio between the price of the table olives and the price of the olive oil...
is established. The ratio in the year $t$ is defined as $r_t^{\text{pr}} = \frac{p_t^{\text{olive}}}{p_t^{\text{transf}}}$
. We get the following ratios: $r_{85}^{\text{pr}} = 0.58$, $r_{86}^{\text{pr}} = 0.26$, $r_{87}^{\text{pr}} = 0.22$, $r_{88}^{\text{pr}} = 0.32$, and $r_{89}^{\text{pr}} = 0.21$. An average of these ratios is calculated, it is equal to 0.32. Then, the price of the olive oil in the year $t$ is equal to $p_t^{\text{olive,est}} = \frac{p_t^{\text{olive}}}{0.32}$ for $t \in [1975, 1984]$, that is, the price of the table olives in the year $t$ over the average ratio.

For the grape to transformation we perform a similar exercise but, in this case, we use as reference crop the table grapes. The price of the grape to transformation is available by varieties and regions but only for three years: 1995, 2000, and 2008.\(^{30}\) We choose a representative variety, tempranillo, and calculate an average price, one per year. Then we compare the prices of the table grapes and the grape to transformation and calculate a ratio for each year. The ratio in the year $b$ is defined as $r_b^{\text{gg}} = \frac{p_b^{\text{grape}}}{p_b^{\text{transf}}}$, $b = 95, 00, 08$. We get the following ratios: $r_{95}^{\text{gg}} = 0.74$, $r_{00}^{\text{gg}} = 0.65$, and $r_{08}^{\text{gg}} = 1.13$. Using these ratios and the prices of the table grapes we estimate a price for the grape to transformation. The price of the grape to transformation in the year $t$ is equal to $p_t^{\text{grape,transf}} = \frac{p_t^{\text{grape}}}{r_t^{\text{gg}}}$ where $b = 95, 00, 08$. The ratio $r_{95}^{\text{gg}}$ is used for the period $t \in [1975, 1996]$, the ratio $r_{00}^{\text{gg}}$ is used for the period $t \in [1997, 2000]$, and the ratio $r_{08}^{\text{gg}}$ is used for the period $t \in [2001, 2011]$.

Finally, we apply an special technique for the citrus. The GAEZ reports information about citrus while the AEA reports more disaggregated information (the group of citrus is composed by oranges, mandarins, and lemons). In order to do data comparable we calculate a weighted average to get a representative values for citrus. For each year $t$ we define the representative price of citrus as

$$p_t^{\text{citrus}} = p_t^o \left( \frac{q_t^o}{q_t^o + q_t^m + q_t^l} \right) + p_t^m \left( \frac{q_t^m}{q_t^o + q_t^m + q_t^l} \right) + p_t^l \left( \frac{q_t^l}{q_t^o + q_t^m + q_t^l} \right),$$

where $o$ refers to oranges, $m$ refers to mandarins, and $l$ refers to lemons, $p_t^i (i = o, m, l)$ is the price in the year $t$ and $q_t^i (i = o, m, l)$ is the national output in year $t$, that is, $q_t^i = \sum_{r=1}^{50} q_{rt}^i$. The weight that each price has in the representative price $p_t^{\text{citrus}}$ depends on the weight that each variety has in the total production. The total value (by region $r$ and year $t$) is calculated as the product of quantities and prices

$$Value_{rt}^{\text{citrus}} = q_{rt}^o p_t^o + q_{rt}^m p_t^m + q_{rt}^l p_t^l.$$

\(^{30}\)Information about prices are extracted from Lissarrague García-Gutiérrez and Martínez de Todo Fernández, page 70, table 31. According to the previous study the price source is: ‘La semana Vitivinícola, Nº 2565 (1995), 2826 (2000), 3240 (2008).’
Finally, a representative production by region $r$ and year $t$ is calculated using the total value and the representative price,

$$q_{citrus}^{rt} = \frac{Value_{citrus}^{rt}}{p_{t}^{citrus}}.$$