Massol, O. & Banal-Estanol, A. (2012). Export diversification and resource-based industrialization: the case of natural gas (Report No. 12/01). London, UK: Department of Economics, City University London.

School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Export Diversification and Resource-based Industrialization: the Case of Natural Gas

Olivier Massol[1]
IFP School
& City University London

Albert Banal-Estañol
City University London
& Universitat Pompeu Fabra

Department of Economics
Discussion Paper Series
No. 12/01

[1] Center for Economics and Management, IFP School, 228-232 av. Napoléon Bonaparte, F-92852 Rueil-Malmaison, France;
Email: olivier.massol@ifpen.fr. Tel.: +33 1 47 52 68 26; fax: +33 1 47 52 70 66.

Export diversification and resource-based industrialization: the case of natural gas ☆

Olivier MASSOL [a,b,*] **Albert BANAL-ESTAÑOL** [b,c]

December 22, 2011

Abstract

For resource-rich economies, primary commodity specialization has often been considered to be detrimental to growth. Accordingly, export diversification policies centered on resource-based industries have long been advocated as effective ways to moderate the large variability of export revenues. This paper discusses the applicability of a mean-variance portfolio approach to design these strategies and proposes some modifications aimed at capturing the key features of resource processing industries (presence of scale economies and investment lumpiness). These modifications help make the approach more plausible for use in resource-rich countries. An application to the case of natural gas is then discussed using data obtained from Monte Carlo simulations of a calibrated empirical model. Lastly, the proposed framework is put to work to evaluate the performances of the diversification strategies implemented in a set of nine gas-rich economies. These results are then used to formulate some policy recommendations.

JEL classification: O13, O14, O2, Q3, Q4

Keywords: Resource-based industrialization; Mean-variance portfolio; Export earnings volatility; Natural gas.

☆ We are grateful to Aitor Ciarreta-Antuñano, Steven Gabriel, Carlos Gutiérrez-Hita, Saqib Jafarey and Jacques Percebois for insightful comments on earlier versions of this paper. We thank Ibrahim Abada, Vincent Brémond, Jeremy Eckhause, Jean-Pierre Indjehagopian and Frédéric Lantz for useful discussions. We also gratefully acknowledge helpful comments received from participants at EURO 2010, INFORMS 2010 and 2011, the 2011 Congreso de la AEEE (Barcelona), the 2011 Transatlatic Infraday (Washington D.C.) and at a seminar organized by the Energy Studies Institute (National University of Singapore). Of course, any remaining errors are ours. The views expressed herein are strictly those of the authors and are not to be construed as representing those of IFP Energies Nouvelles.

a Center for Economics and Management, IFP School, 228-232 av. Napoléon Bonaparte, F-92852 Rueil-Malmaison, France.

b Department of Economics, City University London, Northampton Square, London EC1V 0HB, UK.

c Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain.

* Corresponding author. Tel.: +33 1 47 52 68 26; fax: +33 1 47 52 70 66.
E-mail address: olivier.massol@ifpen.fr

1. Introduction

Export diversification has long been a stated policy goal for many commodity-dependent developing economies. During the last 40 years, many analysts and policy makers have advocated a wave of export-oriented industrialization centered on primary products obtained from resource processing. Their arguments typically emphasize the benefits derived from an increase in the retained value added, or the opportunity to monetize a potentially wasted resource[1] (Pearson, 1970; Hassan, 1975; ESMAP, 1997; MHEB, 2008). Interestingly, natural resources generally offer multiple export-oriented monetization opportunities. In the case discussed in this paper, that of natural gas, methane can be either: exported using transnational pipelines or Liquefied Natural Gas (LNG) vessels, used as a source of power in electricity-intensive activities (e.g. aluminum smelting), converted into liquid automotive fuels, or processed as a raw material for fertilizers, petrochemicals or steel.

This paper aims at assessing the performance of the resource-based export diversification strategies. Again, in the case of natural gas, a wide variety of possible patterns of monetization exist that ranges from one extreme, a whole specialization in raw gas exports as in Yemen, to the other extreme, a total diversification through gas processing industries similar to those experienced in Trinidad & Tobago during the 1980's (Auty and Gelb, 1986). The theoretical basis of our approach stems from the pioneering paper by Brainard and Cooper (1968) that adapts Markowitz's (1952, 1959) Mean–Variance Portfolio (hereafter MVP) theory to analyze the trade-offs between the gains derived from diversification and those resulting from specialization. On the one hand, a wisely selected export diversification may look desirable to moderate the variability of the export earnings of a commodity-dominated economy. But, on the other hand, such a policy can also have a negative and substantial impact on the perceived resource rents if it involves shifting resources from a highly profitable processing industry into substantially less profitable uses.

The main contribution of this paper is to offer a modified MVP model that explicitly takes into consideration the cost structure of these processing industries. Paradoxically, previous studies have usually disregarded processing costs (e.g., Love, 1979; Caceres, 1979; Labys and Lord, 1990; Alwang and Siegel, 1994, Bertinelli et al., 2009).[2] Such an omission seems reasonable in the case of export goods with comparable production costs but can hardly be advocated when processing costs differ significantly as is likely to be the case with resource-based industries.[3] Indeed, any optimal portfolio obtained while focusing solely on export earnings could be largely suboptimal from the perspective of a governmental planner concerned by both the variability of export earnings and the expected amount of resource rents to be perceived. In this paper, we make use of cost information derived from

---

[1] *For example, oil extraction operations in LDCs have historically been associated with extensive flaring or venting of the volumes of natural gas that are jointly extracted with oil.*

[2] *To justify this omission, Bertinelli et al. (2009) underline the unavailability of complete information on the costs of producing one unit of each of the products that could be exported to the world market.*

[3] *In the case of natural gas, the processing cost differs significantly from one type of gas-based industry to another (Auty, 1988a; ESMAP, 1997).*

engineering studies because, despite their inherent limitations, these engineering data reflect the information available to governmental planners (e.g., ESMAP, 1997, 2004). These studies convey some interesting features of the industries under scrutiny, such as an order of magnitude for the economies of scale that can be obtained at the plant level, the ranges of possible capacities for the processing plants...

Our approach can be decomposed into four successive steps. To begin with, we formulate a modified MVP model that embeds an engineering-inspired representation of the resource processing technologies. In a second step, we use natural gas as a case study to address the questions typically faced by practitioners when applying such an MVP-based approach. In particular, this example allows us to detail a careful modeling of the random revenues aimed at being used as an input for the MVP model. Thirdly, the proposed methodology is put to work to examine the optimal export-oriented industrialization strategies that could be implemented in a sample of nine gas-rich countries. Lastly, an adapted gauging methodology is developed to assess the performance of a given export-oriented RBI policy.

We believe that such a tool is valuable for professionals and scholars interested in the design of an export-oriented industrialization policy in a small open economy. It is also of paramount importance for public decision makers in resource-rich countries who have to deal with politically sensitive issues concerning the monetization of the national resources as the obtained results can then be used to formulate some policy recommendations. In the case of natural gas, our findings: (i) confirm that an export-oriented diversification based on resource processing industries is not necessarily a *panacea*, (ii) indicate that some countries should investigate the possibility of rebalancing their current resource monetization strategy, and (iii) question the relevance of certain gas-based industries that have recently received an upsurge in interest. More importantly, these results also indicate that raw exports of natural gas provide the country with the highest level of expected returns, suggesting that any attempts to diversify the economy away from raw export using RBI will involve some trade-offs as such a policy indubitably results in a lowered level of expected returns.

The paper is organized as follows: Section 2 provides a brief overview of the literature related both to Resource-Based Industrialization (hereafter RBI) and to export diversification in the context of a commodity-dominated economy. Section 3 presents a modified MVP model that incorporates an engineering-inspired representation of the resource processing technologies. Section 4 details an application of this methodology to the case of natural gas and clarifies the implementation of the modified model. Then, Section 5 discusses the gas-based industrialization strategies implemented in nine countries with the help of an adapted non-parametric measure of their inefficiencies. Finally, the last section offers a summary and some concluding remarks. For the sake of clarity, all the mathematical proofs are in Appendix A.

2. Literature review

In this section, an overview of the existing literature is provided so as to clarify the context of our analysis. To begin with, the discussion highlights some key aspects related to the volatility of export revenues and its possible influence on the development of a resource-dominated economy. Then, the lessons learned from past RBI experiences are presented. Lastly, we review the application of MVP concepts to evaluate export diversification policies.

2.1 On export volatility, the resource curse and export diversification

Experience provides numerous cases of commodity-dependent economies, particularly countries with a sizeable endowment of hydrocarbons, whose economic performances are outperformed by those of resource-poor economies (Gelb, 1988; Sachs and Warner, 1995; Auty, 2001), a phenomenon coined the "resource curse".[4] What mechanisms might explain this negative relationship between resource abundance and economic performance? Unsurprisingly, several surveys and review articles confirm that this question has motivated a rich literature (Ross, 1999; Stevens, 2003; Frankel, 2010; van der Ploeg, 2011). The proposed explanations can roughly be regrouped in two main categories. A first line of research focuses on governance issues and typically emphasizes the effects of rapacious rent-seeking, of corruption, or those of weakened institutional capacity (Ross, 1999; Isham et al., 2005). A second type of transmission mechanism emphasizes the importance of economic effects such as the controversial Prebisch-Singer thesis of an alleged secular decline in the prices of exported primary commodities relative to those of imported manufactures (Prebisch, 1950; Singer, 1950), or the "Dutch Disease" effect detailed in Corden and Neary (1982).

This latter category also includes the recent explanations based on the volatility of primary commodity prices. The empirical analyses reported in Mendoza (1997), Blattman et al. (2007) and van der Ploeg and Poelhekke (2009) indicate that fluctuations in terms of trade can have a significant negative effect on growth. Several economic arguments may justify these empirical findings. For example, the literature on irreversible investment suggests that the uncertainty associated with this volatility can delay aggregate investment and thus depress growth (Bernanke, 1983; Aizenman and Marion, 1991; Pindyck, 1991). Alternative explanations emphasize either the influence of terms of trade variability on precautionary saving and consumption growth (Mendoza, 1997), or the interactions between trade specialization and financial market imperfections (Hausmann and Rigobon, 2003). Anyway, whatever the exact nature of the mechanism at work, this literature indicates that the variability of natural resource revenues induced by volatile primary commodity prices could be harmful for those economies with the highest concentrations of commodity exports. This perspective provides the motivation of the present study.

---

[4] *However, Alexeev and Conrad (2009) have recently challenged the existence of such a "resource curse" as their empirical findings indicate that oil and mineral wealth have positive effects on income per capita, when controlling for a number of variables.*

Beyond the usual recommendations for a sound macroeconomic management of resource-rich economies, at least three types of strategies have been proposed to handle these volatile revenues. Firstly, the use of market-based financial instruments can make it possible to hedge commodity price risk for a given period of time. This solution has been widely advocated, but Devlin and Lewin (2005) underline that risk management techniques are not so commonly used by the governments of resource-rich countries. A second possible strategy consists of the creation of a dedicated stabilization fund similar to the natural resource fund introduced in, for example, Norway. However, evidence suggests that the effectiveness of funds in mitigating economic volatility is variable depending on country-specific circumstances (Davis et al., 2001) and/or on the details of the fund's institutional procedures (Humphreys and Sandbu, 2007). Lastly, a third strategy echoes the empirical observations of Love (1983, 1986) that document a positive linkage between commodity concentration and export earnings volatility. According to that perspective, countries should consider the implementation of an export diversification policy aimed at moderating the export earnings instability.

The scope of this export diversification must be judiciously selected. Here, we have to distinguish between the case of manufactures and those of processed primary products that are directly derived from the raw commodity. At least, two lines of arguments contest the relevance of a diversification centered on the expansion of manufactured exports. Firstly, the "Dutch Disease" effect (Corden and Neary, 1982) may compromise the chances of a successful wave of export-oriented industrialization centered on manufactured goods. Secondly, the empirical findings of Love (1983) suggest that a broad diversification into manufactures does not necessarily lead to greater earnings stability for a commodity-dominated economy. On the contrary, primary processing could constitute an attractive move. For example: Owens and Wood (1997) build on the Heckscher-Ohlin (H-O) trade theory and indicate that resource-rich countries with a moderate to large endowment in skilled labor can have a comparative advantage in processed primary goods. That's why we have decided to narrow our analysis down to the case of an export diversification centered on the installation of resource-based industries.

2.2 Resource-based industrialization, a review

From time to time, it has been emphasized that RBI could also constitute an appropriate medium to achieve the "big push" advocated by Rosenstein-Rodan (1943), i.e., a large stimulus able to catapult an economy away from a low-level equilibrium trap. However, experience earned with over-ambitious RBI policies has generally provided mixed or disappointing results, as illustrated in the collection of cases presented in Auty (1990). There have been a number of explanations which analyze the causes of these unsatisfactory results. For example, the literature in the tradition of Hirschman (1958) stresses that RBI is unlikely to stimulate growth in the rest of the economy, particularly if this sector is dominated by foreign firms that are allowed to repatriate their profits, because this sector would produce few powerful forward and backward linkages to other sectors. In a survey, Roemer (1979) rejects the "one size fits all" arguments in favor of RBI and calls for case-specific industrial strategies that should take into consideration the nature of the country's comparative advantages, and the

specific features of the industrial organization. Besides, it seems that detailed implementation mechanisms matters. Looking at past RBI experiences in petroleum-exporting countries, Auty (1988b) notices that state-owned enterprises played a key role in the implementation of RBI policies and documents the influence of these firm's governance and management on the observed performances of the RBI policies. In this paper, we do not discuss these important governance issues but concentrate our attention on the identification of an optimal RBI policy.

In the sequel of this paper, we assume that the countries under scrutiny have an appropriate endowment in skilled labor and that an efficient governance structure in the industrial sector can be implemented.

2.3 Export diversification, an MVP analysis

Is export diversification suitable or not? If yes, an interesting question emerges: which products should be given priority over others? To answer these questions, Brainard and Cooper (1968) proposed applying the MVP concepts developed in Markowitz (1952, 1959). Originally, MVP was intended to analyze the optimal composition of a portfolio of financial securities, though numerous applications rooted in a non-financial context have been proposed over the years.[5] For the sake of brevity, we do not review the vast literature related to MVP, instead we concentrate our discussion on papers connected to Brainard and Cooper (1968).

The MVP approach has been widely applied to the analysis of exports earnings (e.g., Love, 1979; Caceres, 1979; Labys and Lord, 1990; Alwang and Siegel, 1994; Bertinelli et al., 2009). In these contributions, the authors are concerned with the export decisions of a given country. Commodity prices are assumed to be the unique source of uncertainty and these random variables are supposed to be jointly normally distributed with known parameters (i.e., the vector of expected values and the variance-covariance matrix). The decision variables are the shares of the various products in the country's total exports and together constitute the country's export portfolio. The country's utility to be maximized is modeled using a mean-variance utility function that captures the trade-offs between the risks measured by the portfolio's variance and returns measured by the expected amount of export earnings. Additionally, the chosen portfolio must be a feasible one. Hence, the country's optimization program is subject to constraints aimed at describing the set of feasible export combinations. This analytical framework is thus completely equivalent to the typical MVP model with no riskless asset and no short sales permitted. From a computational perspective, it can be formulated as a quadratic programming problem. By continuously varying the coefficient of absolute risk aversion, it is possible to determine a set of optimal portfolios and draw an efficient frontier in the plane (variance of the country's export earnings, expected value of these export earnings).

[5] *A recent industrial example is given by Roques et al. (2008) who, in the context of energy planning, have applied the MVP approach to analyze technology choices in liberalized energy markets.*

3. RBI-based export diversification, a portfolio approach

In this section, we present the problem faced by the governmental planners of a resource-rich country that seeks to implement an export strategy focusing on RBI.

3.1 The basic framework

We consider the risk-averse government of a small open economy endowed with a unique resource[6] and examine the government's export-oriented options to monetize that resource. In most countries, the government claims an ultimate legal title to the nation's resources, even those located in a private domain.[7] It can grant users rights as concessions if it so chooses. Nonetheless, it remains the exclusive or almost exclusive recipient of the resource rents and has thus a considerable influence on the monetization of the country's domestic resources. Potentially, there are m exported goods produced domestically and derived from the processing of the country's resource. Hence, we assume that the influences of the other non-resource-based exports can be neglected so that attention can be entirely focused on the export earnings generated by these m resource-based industries. There are no joint products in these resource-processing industries.

The government's decision amounts to choosing a resource monetization policy for a given planning horizon, i.e., the flows of products exported during the planning horizon. We assume that the monetization strategy selected at the beginning of this planning horizon is held unchanged to the terminal date. This assumption is coherent with the irreversible nature of the capital investments required for the implementation of a resource-processing industry. During the planning horizon, the instantaneous flow of resource aimed at being either exported or processed is constant and known. This simplifying assumption could easily be relaxed to deal with a known, but unsteady, pattern of resource flow during the planning horizon.

The country in question is small and is a price taker in the sense that it is unable to influence the international prices. This assumption seems appropriate for numerous resources and their associated processed primary products. It has also been used in numerous applied studies (Love, 1979; Labys and Lord, 1990; Alwang and Siegel, 1994; Bertinelli et al., 2009). The government makes its economic decisions before international prices are known. We assume away other types of uncertainty. Hence, our analysis concentrates on price risk and does not consider other technical or operational risks (e.g., through domestic input price, plant outages, construction cost overruns...). Given that domestic conditions are usually better known, it seems reasonable to assume that foreign prices are less likely to be known with certainty. The international prices of the exported goods are assumed to be jointly normally distributed.

---

[6] *The extension to the more general case of more than two unrelated resources (i.e., resources that can be processed using industries that have no more than one resource in their list of inputs) does not cause any conceptual difficulty.*

[7] *This institutional framework is very common for underground resources (both mineral and petroleum). It can also be occasionally observed with above ground resources (a famous example is provided by the case of hydropower resources in Norway which are tightly controlled by the state).*

3.2 Taking processing costs into account

The main contribution of this paper is to include processing costs derived from engineering studies.

For each exported good i, the governmental planners have to decide on an industrial configuration i.e., the number of plants n_i to be installed and $\left(q_{ij}\right)_{j\in\{1,...,n_i\}}$ the non-negative resource flows aimed at being processed in these various plants. Considering the m exported goods together, we let $n=\left(n_1,...,n_i,...,n_m\right)^T$ describe the number of plants installed for each exported goods, and use $\left(q\right)_n$ as a short notation for the whole list of resource-processing decisions $\left(q_{ij}\right)_{i\in\{1,...,m\},\, j\in\{1,...,n_i\}}$. We also denote $q=\left(q_1,...,q_i,...,q_m\right)^T$ where q_i is the total resource flow transformed into good i, i.e. $q_i=\sum_{j=1}^{n_i} q_{ij}$.

We can now detail the cost of each resource-processing industry. For an individual plant $j\in\{1,...,n_i\}$, we denote: y_{ij} the output, q_{ij} the amount of resource used as an input and x_{ij} a vector that gathers all the other inputs (capital, labor, other intermediate materials...). The resource input q_{ij} and all the combinations of the other inputs x_{ij} are assumed to be perfect complements.[8] Thus, the productivity of the resource input y_{ij}/q_{ij} is equal to a constant positive coefficient a_i that is invariant with the activity level y_{ij}. Using this linear relation, the plant's cost function can be reformulated as a single-variable function of the resource input q_{ij}. The total cost of installing and operating a plant capable of processing any given flow of resource $q_{ij}\in\left[\underline{Q_i},\overline{Q_i}\right]$ is $c_i\left(q_{ij}\right)$ where $c_i\left(.\right)$ is a positive, monotonically increasing, twice continuously differentiable concave cost function of the variable q_{ij}. Because of technological constraints on the feasible combinations of the other inputs x_{ij}, some lumpiness is at work at the plant level and the plant's cost function is defined on the exogenously restricted domain $\left[\underline{Q_i},\overline{Q_i}\right]$, where $\underline{Q_i}$ (respectively $\overline{Q_i}$) is the plant's minimum (respectively maximum) implementable size. This interval is large enough to verify $0<2\underline{Q_i}\leq\overline{Q_i}$. If the output were to be null, there is no need to build a plant and we impose that $c_i\left(0\right)=0$.

---

[8] *Hence, we are implicitly assuming that all the other inputs are as a group separable from the resource input so that the plant's production function has the following nested form:* $y_{ij}=a_i\ Min\left(q_{ij},k_i\left(x_{ij}\right)\right)$ *where the first stage corresponds to a Leontief fixed proportion technology, and the second stage is described by an intermediate production function* k_i *that is assumed to be well-behaved (i.e. positive, monotonic, twice continuously differentiable and quasi-concave). The resource input and the bundle* $k_i\left(x_{ij}\right)$ *are used in fixed constant proportions are thus perfect complements.*

We assume that the country's total production cost function is additively separable in the processing technologies. That is,

$$C\left(n,(q)_n\right)=\sum_{i=1}^{m}\sum_{j=1}^{n_j}c_i\left(q_{ij}\right), \tag{1}$$

where $C\left(n,(q)_n\right)$ is the country's total production cost of the industrial configurations generated by n and $(q)_n$.

The government has a constant absolute risk aversion utility which, coupled with the normal assumption above, leads to a mean-variance utility function. Thus, we are assuming that export decisions can be derived from the following aggregate utility maximization problem:

Problem (P0) max $$U\left(n,(q)_n\right)=\overline{R}^T q - C\left(n,(q)_n\right) - \frac{\lambda}{2}\, q^T\Phi\, q, \tag{2}$$

s.t. $$\sum_{i=1}^{m}\sum_{j=1}^{n_i}q_{ij} = PROD, \tag{3}$$

$$q_{ij}\in\{0\}\cup\left[\underline{Q_i},\overline{Q_i}\right],\quad \forall i\in\{1,...,m\},\ \forall j\in\{1,...,n_i\} \tag{4}$$

$$n\in\left\{\mathbb{N}^*\right\}^m \tag{5}$$

where $\overline{R}=\left(\overline{R_i}\right)^T$ is the vector of expected unit revenues, Φ the associated variance-covariance matrix, λ is the coefficient of absolute risk aversion, and $PROD$ the overall flow of resource aimed at being processed for export.

The objective function (2) captures the trade-offs between the gains in terms of reduction in export earning instability and the gains in terms of increase in the expected value of the perceived resource rents. Equation (3) is the resource constraint and the constraints of type (4) and (5) together represent the lumpiness of the resource-processing technologies.

We now discuss how the restrictions associated to the lumpy nature of the processing technologies at hand impact the MVP problem. From equations (3) and (4), we can remark that the problem has no solution if $PROD$ is strictly less than $\mathrm{Min}_i\left\{\underline{Q_i}\right\}$. So, RBI cannot be encouraged unless the overall flow of resource aimed at being processed for export is large enough to justify at least the construction of a resource-processing plant with the smallest implementable size.[9] For larger values of the overall flow of resource, one may also wonder if the existence of an industrial configuration is necessarily granted as lumpiness could preclude the satisfaction of the resource equation (3). The following

[9] *This situation offers some resemblances with Perold (1984) that discusses the case of a financial portfolio manager who seeks to prevent the holding of very small active positions (because small holdings usually involve substantial holding costs while offering a limited impact on the overall performance of the portfolio).*

lemma addresses this concern and shows that this should not be a problem as there systematically exists at least one industrial configuration capable of processing any flow of resource larger than $\text{Min}_i \left\{\underline{Q_i}\right\}$.

Lemma 1: *Suppose that the plant's maximum and minimum sizes satisfy* $0 < 2\underline{Q_i} \leq \overline{Q_i}$, $\forall i = 1,...,m$. *For any flow of resource* $PROD$ *with* $PROD \geq \text{Min}_i \left\{\underline{Q_i}\right\}$, *there exists at least one industrial configuration, i.e.* n *a vector of* m *integers and* $(q)_n$ *the resource flows processed in the various plants, that jointly satisfies the conditions stated in equations* (3), (4) *and* (5).

3.3 A computationally tractable formulation

From a computational perspective, the formulation used in problem (P0) makes it relatively awkward to manipulate because the total number of real-valued variables q_{ij} is given by n the vector of integer variables. In this subsection, we propose a reformulation aimed at making this problem more tractable by taking advantage of the problem's specificities.

Our approach is based on the following remark: for a given set of exogenous parameters and some given level of export for each good (thus a given vector q), there may exist many industrial configurations that jointly satisfy the problem's constraints (3), (4) and (5). By construction, all these configurations offer the same level of expected total revenues $\overline{R}^T q$ and the same total risk $q^T \Phi\, q$, but do not have the same processing cost. According to the objective function, some of these configurations should be preferred to others: those that minimize these processing costs. Now, we provide a characterization of a cost-minimizing industrial configuration.

In the following proposition, we focus on a given good i and provide, for any flow of resource q_i larger than $\underline{Q_i}$, the composition of the cost-minimizing industrial configuration capable of processing exactly that flow. Hereafter, we denote $n_i(q_i)$ the smallest number of plants that can be installed to transform a given flow q_i i.e. $n_i(q_i) = \lceil q_i / \overline{Q_i} \rceil$ where $\lceil . \rceil$ is the ceiling function. We also denote $r_i(q_i) = q_i - (n_i(q_i) - 1)\overline{Q_i}$ that would measure the size of the residual plant if $n_i(q_i) - 1$ plants were to be installed with the largest implementable size.

Proposition 1: *Suppose: that the plant's maximum and minimum sizes satisfy* $0 < 2\underline{Q_i} \leq \overline{Q_i}$, $\forall i = 1,...,m$ *and that the plant level cost functions* $c_i(x_i)$, $\forall i = 1,...,m$ *satisfy the assumptions above (concavity, continuity, differentiability). For any flow of resource* q_i *aimed at being transformed into good* i, *with* $q_i \geq \underline{Q_i}$, *a cost-minimizing*

industrial configuration for that particular good has $n_i(q_i)$ *plants, and each of these plants are processing respectively the following flows of resource:*

- *if* $r_i(q_i) \geq \underline{Q_i}$ *:* $\left(\overline{Q_i},...,\overline{Q_i},...,\overline{Q_i},r_i(q_i)\right)$,

- *otherwise if* $r_i(q_i) < \underline{Q_i}$ *:* $\left(\overline{Q_i},...,\overline{Q_i},\left[r_i(q_i)+\overline{Q_i}-\underline{Q_i}\right],\underline{Q_i}\right)$.

If we denote $\delta_i(q_i)$ an indicator function that takes the value 1 if $r_i(q_i) \geq \underline{Q_i}$ and 0 elsewhere, this proposition can be used to define $C_i\left(q_i,n_i(q_i),\delta_i(q_i)\right)$ the function that gives the minimum total cost to transform any flow of resource q_i using the industry i using the following function:

$$C_i(q_i,n_i,\delta_i)=\begin{cases}(n_i-2).c_i\left(\overline{Q_i}\right)+\delta_i.\left[c_i\left(\overline{Q_i}\right)+c_i\left(q_i-(n_i-1).\overline{Q_i}\right)\right]\\ \quad+(1-\delta_i).\left[c_i\left(\underline{Q_i}\right)+c_i\left(q_i-\underline{Q_i}-(n_i-2).\overline{Q_i}\right)\right].\end{cases} \tag{6}$$

More importantly, this proposition suggests a simplification of the problem (P0). Rather than using the individual plant's inputs q_{ij} as decision variables, we can use the total flow of resources q_i aimed at being transformed in each good i together with the structure of the cost-minimizing industrial configurations provided in Proposition 1. As a result, we now propose a revised specification of the problem (P0):

Problem (P1) max $\sum_{i=1}^{m}\left[\overline{R_i}q_i-\varsigma_i\, C_i(q_i,n_i,\delta_i)\right]-\frac{\lambda}{2}\sum_{i=1}^{m}\sum_{i'=1}^{m}q_i\Phi_{ii'}q_{i'}$ (7)

s.t. $\sum_{i=1}^{m}q_i \;=\; PROD$ (8)

$(n_i-1)\overline{Q_i} \;\leq\; q_i \;\leq\; n_i\overline{Q_i}$ $\forall i\in\{1,...,m\}$ (9)

$(n_i-1)\overline{Q_i}+\underline{Q_i}\;\delta_i \;\leq\; q_i$ $\forall i\in\{1,...,m\}$ (10)

$q_i \;\leq\; \delta_i\overline{Q_i}+(n_i-1)\overline{Q_i}+\underline{Q_i}$ $\forall i\in\{1,...,m\}$ (11)

$q_i \;\leq\; \varsigma_i\; PROD$ $\forall i\in\{1,...,m\}$ (12)

$\underline{Q_i}\;\varsigma_i \leq\; q_i$ $\forall i\in\{1,...,m\}$ (13)

$q_i \geq 0$, $n_i\in\mathbb{N}^*$, $\delta_i\in\{0,1\}$, $\varsigma_i\in\{0,1\}$ $\forall i\in\{1,...,m\}$ (14)

where, for any good i, the decision variables are as follows: the non-negative flow of resource q_i transformed in i; a binary variable ς_i associated with the disjunctive choice "*export at least a certain amount, or not at all*"; the number of plants n_i to be implemented; and δ_i a binary variable that indicates whether $r_i(q_i)$ is larger than $\underline{Q_i}$ or not. In this program, the objective is to maximize the value of the mean-variance utility (7) and this optimization program is subject to a series of linear constraints aimed at describing the set of possible export combinations. Here, (8) is the resource constraint. Thanks to constraints of type (9), n_i has to be related to q_i so that $n_i = \lceil q_i / \overline{Q_i} \rceil$ for any good i. The constraints (10) and (11) together insure that the binary variable δ_i takes the value 1 if and only if $r_i(q_i) \geq \underline{Q_i}$. The constraints (12) and (13) are the minimal transaction unit constraints proposed in Perold (1984) and jointly create the conditions for a disjunctive choice: constraint of type (12) forces the binary variable ς_i to be equal to 1 if the country wishes to process a strictly positive flow of resource using that industry (i.e., $q_i > 0$) while constraint (13) imposes the processing of at least a certain volume $\underline{Q_i}$ in case of a strictly positive volume q_i. Thus, export of product i will be impeded if the desired flow q_i is strictly less than the prescribed minimum size $\underline{Q_i}$.

Proposition 2: *Suppose that: (i) the plant level cost functions* $c_i(x_i)$, $\forall i = 1,...,m$ *satisfy the assumptions above (concavity, continuity, differentiability), (ii) that the plant's maximum and minimum sizes satisfy* $0 < 2\underline{Q_i} \leq \overline{Q_i}$, $\forall i = 1,...,m$, *and (iii) that the overall flow of resource is large enough to justify at least the construction of a plant i.e.* $PROD \geq \text{Min}_i \{\underline{Q_i}\}$. *In that case, the problem (P1) has a global solution.*

More compactly, we can simply note that the problem (P1) is a single-period mean-variance portfolio problem under separable concave transaction costs with minimal share constraints and integer constraints on the continuous variables. From a computational perspective, the formulation used in problem (P1) seems preferable because the number of non-negative variables is restricted to m.

4. Application and policy performance appraisal

In this section, we detail an application of the proposed methodology to assess the performances of the export-oriented industrialization possibilities offered by natural gas.

4.1 Background and data

We aim at analyzing the gas monetization strategies implemented in a sample of nine economies endowed with significant reserves of natural gas (Angola, Bahrain, Brunei, Equatorial Guinea,

Nigeria, Oman, Qatar, Trinidad & Tobago, and the UAE). The gas-processing industries implemented in these countries are overwhelmingly export-oriented.

In this study, we focus on six resource-based industries that represent the major monetization options offered by natural gas and neglect the influences of other exports. The list includes: (i) the liquefaction train (a dedicated cryogenic infrastructure used to export natural gas in an LNG form); metal processing industries like (ii) aluminum smelting or (iii) iron and steel plants producing Direct Reduced Iron (DRI); petrochemical plants converting natural gas into (iv) diesel oil (using the so-called Gas-To-Liquid (GTL) techniques) or (v) methanol (a basic non-oil petrochemical); and (vi) fertilizer industries producing urea.

Both gas fields and gas-based industries are vertically-related, specialized assets in the sense of Klein et al. (1978). Accordingly, investment in these assets generates appropriable quasi rents and creates the possibility of opportunistic behavior in case of separate ownership. Against this backdrop, the transactions involving gas producers and gas processing industries are usually governed by long-term contracts with a very long duration that include binding "take or pay" clauses aimed at tightly limiting the variability of the purchased gas flow. Because of these contractual features, it is sensible to neglect volume variability and to assume, as in our modified MVP model, that the variability of the export revenues is caused entirely by price uncertainty.

Table 1. The size and composition of the planned portfolios

	Gas flow *PROD* (MMCFD)	Allocated Shares (%)						HHI
		Aluminum smelters	GTL plants	DRI plants	LNG trains	Methanol plants	Urea plants	
Angola	938.4	16.0%	-	-	84.0%	-	-	73.2%
Bahrain	342.5	63.5%	-	14.6%	-	11.1%	10.7%	44.9%
Brunei	1 165.8	-	-	-	93.7%	6.3%	-	88.1%
Equatorial Guinea	656.8	-	-	-	85.4%	14.6%	-	75.1%
Nigeria	3 582.6	1.3%	8.9%	-	88.9%	-	0.8%	79.8%
Oman	2 016.2	4.5%	-	2.5%	83.5%	2.4%	7.1%	70.5%
Qatar	12 722.6	1.1%	13.3%	0.6%	82.8%	0.6%	1.5%	70.4%
Trinidad & Tobago	3 069.6	-	0.7%	4.7%	74.6%	18.7%	1.4%	59.4%
U.A.E.	1 454.0	29.0%	-	7.9%	53.7%	-	9.4%	38.8%

Note: For each country, this table details: (i) the overall flow of natural gas used as an input in these six processing industries measured in millions of cubic feet per day (MMCFD), (ii) the shares of this flow allocated to these industries, and (iii) the associated Herfindahl-Hirschman Index. The overall flow is those required for the operation of all the country's gas processing plants at their designed capacities. It has been obtained using the gas input values given in Table B-1 (cf. Appendix B) together with a detailed inventory of the projected output capacities (in tons of output) for the processing plants already installed, those under-construction and the projects for which a "Final Investment Decision" was formally announced as of January 1st 2011. These inventories have been obtained from the US Geological Survey, IHS Global Insight, governments and project promoters.

Table 1 summarizes the gas monetization strategies implemented in these countries, namely (i) the overall flow of natural gas aimed at being processed in these six export industries, and (ii) the composition of the country's portfolio. In addition, a quantitative measure of diversity may usefully provide an overall picture of the implemented portfolio and thus ease cross-country comparisons.

Because of its simplicity, the Herfindahl-Hirschman Index (hereafter HHI), defined as the sum of the squared shares, constitutes an attractive choice. Indeed, the HHI reflects both variety (i.e. the number of industries in operation) and balance (the spread among these industries).

According to Table 1, the overall gas flows to be processed differ a lot from one country to another but remain modest compared to the world gas production that attained 308,962 MMcfd in 2010 (BP, 2011). We can notice that export diversification is at work in all these countries as all of them have implemented at least two industries. Looking at the HHI scores, one may notice that the two most diversified portfolios are those implemented in the UAE and Bahrain. Interestingly, Bahrain is the only country that does not export LNG (i.e., natural gas in a liquefied form) and has thus implemented a complete diversification away from raw exports. On the contrary, a significant share is allocated to LNG export facilities in all the other eight countries. In seven countries, the LNG share is around or above 75% and this preponderance largely explains their high HHI scores.

4.2 Numerical hypotheses

We now detail and discuss the numerical assumptions used in our analysis.

a - Planning horizon

To begin with, we clarify the chronology. Gas-based industrialization typically entails the installation of capital intensive industries. As the corresponding investment expenditures are largely irreversible, planners have to consider an appropriately long planning horizon. We thus follow ESMAP (1997) and consider a construction time lag measured from the moment of the actual start of construction of three years followed by 25 years of operations (this latter figure is supposed to be equal to a plant's entire life-time).

b - Resource extraction

In this study, the stream of future gas extraction is assumed to be imposed by exogenous geological considerations. For a given country, the flow of natural gas that will be extracted during the whole planning horizon is assumed to be known and to remain equal to $PROD$ during that horizon.[10] For each of the countries under scrutiny, we have used the flow figures listed in Table 1.

Here, the country's total extraction cost is a given that does not vary with the composition of the portfolio. Given that publicly available data on E&P costs are rather scarce (these costs vary greatly by region, by field and scale) compared to those available on gas processing technologies, E&P costs have been excluded from the analysis. That's why we have adopted the "netback value" approach that is commonly used in the gas industry.[11] The netback value overestimates the amount of resource rent because the E&P costs have not been deducted. However, adopting either a resource rent perspective,

[10] *Of course, a more complex extraction profile could be considered if appropriate data were available. Nevertheless, this so-called "plateau" profile is very common in the natural gas industry.*

[11] *The netback value per unit volume of gas is defined as the difference between discounted export revenues and discounted processing and shipping costs (Auty, 1988a) and is often interpreted as a residual payment to gas at wellhead.*

or a netback one for the objective function used in our MVP model has no impact on the composition of the optimal portfolios.

c - Processing technologies

The sizes of the individual plants can be continuously drawn within the ranges listed in Table B-1 (cf. Appendix B). We can remark that, for each technology i, the condition $0 < 2\underline{Q_i} \leq \overline{Q_i}$ holds.

Concerning processing costs, project engineers typically evaluate a plant's total investment expenditure using a smoothly increasing function. The specification $c_i\left(q_{ij}\right) = \alpha_i . q_{ij}^{\beta_i}$, where q_{ij} is the processing capacity of plant j and β_i represents the (non-negative) constant elasticity of the total investment cost with respect to production, is a popular choice. With the gas processing technologies at hand, plant-specific economies of scale are at work. Hence, $\beta_i \leq 1$ for all i. In addition, maintenance and operating (O&M) costs are assumed to vary linearly with output. This specification of the plant level cost functions is thus compatible with our modeling framework. From a numerical perspective, all the results presented hereafter are derived from the figures listed in Table B-1 (Appendix B). In this study, common technologies and cost parameters have been assumed for all countries, which is consistent with the method usually applied in preliminary cost estimations of resource processing projects (e.g. ESMAP, 1997).

To our knowledge, there is no foolproof way of choosing the discount rate for such a problem. Here, a 10% figure is assumed. That figure seems reasonable as a current cost of capital in competitive markets, after inflation has been subtracted out. Sensitivity analyses of the results to both a lower (8%) and a higher (12%) cost of capital have also been carried out but did not sensibly modify the conclusions. For the sake of brevity, these results are not reported hereafter.

d - Revenues

Any application of our MVP approach requires some information on the joint distribution of the random revenues. To our knowledge, most past studies use the descriptive statistics computed from world market price series as inputs (Brainard and Cooper, 1968; Labys and Lord, 1990; Alwang and Siegel, 1994; Bertinelli et al., 2009). Accordingly, international prices are supposed to follow a strictly stationary process and the average prices and the estimated variance-covariance matrix are directly used as proxies for the true, but unobserved, values of the expected value and the variance-covariance matrix.

However, two caveats must be mentioned. Firstly, serial correlation is frequently observed in individual commodity price series. As a result, we follow the recommendations stated in Geman (2005, p. 51) and look for an empirical model capable of generating price trajectories that are consistent with the observed dynamics. Secondly, Pindyck and Rotemberg (1990) document the tendency of the prices of seemingly unrelated commodities to exhibit some excess co-movements even

after accounting for macroeconomic effects. They argue that herd effect and liquidity constraints may explain this finding. Their empirical findings have been contested (Leyborne et al., 1994; Deb et al., 1996; Ai et al., 2006), yet one of the great merits of this debate is that it has become apparent that co-movements may possibly be at work between unrelated commodities. In this study, the commodities at hand are clearly related[12] and their price trajectories are likely to exhibit some significant co-movements. As a consequence, this empirical model should also capture the intricate dynamic interdependences among these prices.

A parsimonious multivariate time-series model of the monthly commodity prices has thus been specified and estimated. The construction of this empirical model is detailed in Appendix C. Monte Carlo simulations of this empirical model allow us to generate a large number (100,000) of possible future monthly price trajectories (evaluated in constant US dollars per ton of exported product). These trajectories are used in combination with a Discounted Cash Flow (DCF) model based on the assumptions detailed in Table B-1 (gas input values, conversion factors, cost of raw minerals for aluminum smelting and iron ore reduction) to derive a sample of present values of the revenues obtained when processing one unit of resource with the six industries at hand. This sample is in turn used to estimate the parameters of the multivariate distribution of these present values: the expected value $\overline{R}$ and the variance-covariance matrix Φ.

4.3 The efficient frontier

All these data on both revenues (the estimated parameters $\overline{R}$ and Φ) and costs are used as inputs in our modified MVP model (Problem P1). Hence, we can identify the optimal portfolios of the gas processing technologies for a country that considers a given value for the coefficient of absolute risk aversion.

From a computational perspective, we can notice that: (i) the number of gas-based industries under consideration remains limited ($m = 6$), and (ii) the maximum implementable sizes of the gas-processing plants are large enough to process a significant share of any country's gas production.[13] As a result, the size of the mixed-integer nonlinear optimization problem (MINLP) at hand remains small enough to be successfully attacked by modern global solvers such as BARON (Sahinidis, 1996; Tawarmalani and Sahinidis, 2004). Thanks to recent developments in deterministic global optimization algorithms (branch and bound algorithms based on outer-approximation schemes of the original non-convex MINLPs, range reduction techniques, and appropriate branching strategies), an accurate global solution for this problem can be obtained in modest computational time.

---

[12] *Numerous linkages exist among these commodities. For example: natural gas is a major input into the production of urea or methanol. Natural gas and oil are co-products in numerous cases and gas prices are also notoriously influenced by the oil products indexed pricing formulas used in numerous long-term importing contracts. Both aluminum smelting and steel production are well known energy intensive activities. Besides, these two mineral commodities can be considered as imperfect substitutes in numerous end uses.*

[13] *It is sufficient to compare the values of: (i) gas flows listed in Table 1 and (ii) the maximum processing capacities listed in Table B-1 (cf. Appendix B).*

By varying the coefficient of absolute risk aversion, it is possible to determine the efficient frontier, i.e., the set of feasible optimal portfolios whose expected returns (i.e. the expected present values of future export earnings net of processing costs) may not increase unless their risks (i.e., their variances) increase. Hence, this approach does not prescribe a single optimal portfolio combination, but rather a set of efficient choices, represented by the efficient frontier in the graph of portfolio expected return against portfolio standard deviation. Depending on the country's own preferences and risk aversion, planners can choose an optimal portfolio (and thus a risk-return combination).

Figure 1. The efficient frontier, an illustration for Bahrain and the UAE

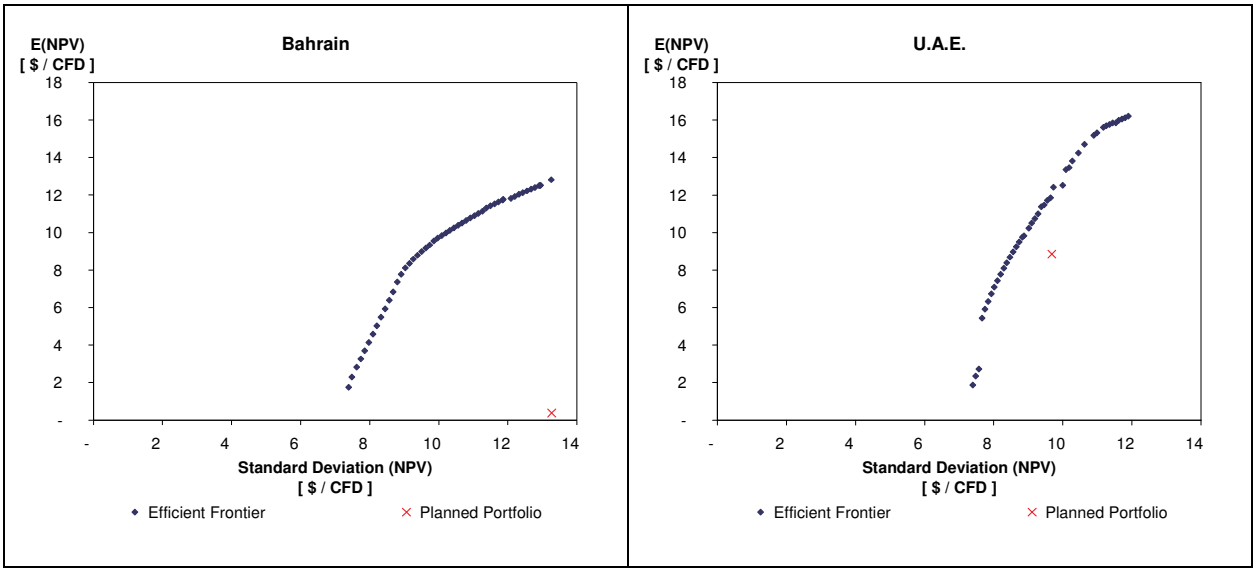


Figure 1 shows the obtained efficient frontier. From that figure, several facts stand out. First, the efficient frontier illustrates the presence of trade-offs between risk and reward: the higher returns are obtained at a price of a larger variance. This figure also confirms that RBI-based export diversification policies cannot totally annihilate the commodity price risks as the total risk associated with minimal risk portfolio remains strictly positive.

Second, we can notice that, contrary to the frontier obtained using the standard MVP formulation, the efficient frontiers at hand exhibit some discontinuities. Given that the modified MVP model includes some binary/integer variables, continuously varying the coefficient of absolute risk aversion from a given value to a neighboring one may cause the model to switch from an initial optimal industrial configuration (described by a combination of binary and integers) to another one that can be quite different in terms of processing costs.

Third, we can compare these frontiers. For low levels of risks, the expected returns are similar. For these points we have noted that the composition of the efficient portfolios is similar (a combination of mineral processing activities: aluminum smelting and iron ore processing). On the contrary, for large enough levels of risk (a standard deviation larger than 10 $/cfd), UAE's efficient portfolios obtain a larger expected return than Bahrain's one. This plot-inspired remark calls for a closer investigation of the composition of the efficient portfolios. In fact, the main difference is connected with the countries'

production levels. Interestingly, if we evaluate, for each technology, the range of the expected net present value of the export earnings net of processing costs in $/cfd as a function of the plant's size, we find that raw gas exports based on the LNG technology systematically provide the largest returns. Because of this absolute domination of LNG exports, the greater the appetite for returns of planners, the more LNG plants there would be in the optimal portfolio. But, a full specialization in the export of LNG is not necessarily feasible because of lumpiness issues. Indeed, a comparison of the minimum implementable sizes (measured in terms of resource flows requirements) indicates that LNG export facilities have a very large-scale nature compared to alternative monetization options. So, the LNG option is only implementable in countries with sufficiently large resource endowments, which is not the case for Bahrain. Incidentally, the fact that LNG provides the largest returns explains why risk-neutral project promoters generally perceive this option to be the most attractive.[14] As a corollary, we can note that: for a country with a specialized export structure fully concentrated on LNG (i.e., on raw exports of natural gas), any attempt to diversify will involve some trade-offs: a lower risk will be obtained at a price of a smaller return...

Table 2. The performances of the implemented portfolios q_0

	Expected return E_0 ($/CFD)	Standard Deviation $\sqrt{V_0}$ ($/CFD)
Angola	12.98	10.19
Bahrain	0.38	13.27
Brunei	15.84	11.56
Equatorial Guinea	15.53	11.21
Nigeria	14.78	11.85
Oman	14.49	11.13
Qatar	14.13	11.87
Trinidad & Tobago	14.86	10.86
U.A.E.	8.85	9.68

Note: For each country, the export policy q_0 reported in Table 1 has been used to evaluate both the expected present value E_0 and the associated variance V_0 of future export earnings (net of processing costs). Concerning E_0, the cost function suggested in Proposition 1 has been assumed. Hence, $E_0 = \sum_{i=1}^{m} \left[\overline{R_i} q_{0i} - \varsigma_{0i} \; C_i \left(q_{0i}, n_{0i}, \delta_{0i} \right) \right]$ where $n_{0i} = \left\lceil q_{0i} \big/ \overline{Q_i} \right\rceil$, ς_{0i} is a binary variable that takes the value 1 if $q_{0i} > 0$, and δ_{0i} is a binary variable that takes the value 1 if $r_i \left(q_{0i} \right) \geq \underline{Q_i}$. Concerning the risks, the reported variance is $V_0 = q_0^{\,T} \Phi \; q_0$.

Last but not least, this figure can also be used to appraise the efficiency of the gas monetization policy $q_0 = \left(q_{01}, ..., q_{0m} \right)^T$ chosen by the governmental planners (i.e., those detailed in Table 1). Indeed, the numerical hypotheses above can be used to evaluate both the expected return E_0 and the risk V_0 of a

[14] *As an illustration, we can quote the case of Yemen where LNG exports started in 2009 and those of Papua New Guinea where two major LNG projects are actively promoted by international petroleum companies.*

particular portfolio q_0. In Table 2, we report the figures obtained for the nine countries under scrutiny. In Figure 1, the countries' efficient frontiers are graphed together with a point representing the performance of the country's portfolio q_0 in terms of risks and returns. So, a simple visual evaluation of distance from the efficient frontier provides a useful indication of the inefficiencies resulting from the chosen diversification policy.

5. Portfolio efficiency appraisal

To complete the visual indications above, we now provide a quantitative evaluation of the efficiency of the planned portfolio.

5.1 Methodology

We use an adapted version of the non-parametric portfolio rating approach proposed in Morey and Morey (1999) and further generalized in Briec et al. (2004). According to this approach, the inefficiency of a given portfolio is evaluated by looking at the distance between that particular element in the production possibility set and the efficient frontier.

Formally, we analyze the case of a country that considers a feasible[15] gas monetization policy $q_0 = (q_{01}, ..., q_{0m})^T$ that has a given level of expected return E_0 and a given risk V_0. Starting from this portfolio with unknown efficiency, we apply a directional distance function that seeks to increase the portfolio's expected net present value while simultaneously reducing its risk. If we consider the direction given by the particular vector $g = (-g_V, g_E) \in (-\mathbb{R}_+) \times \mathbb{R}_+$, this distance is given by the solution of the following MINLP:

$$\text{Problem (P2)} \quad \max \quad \theta \tag{15}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} \left[\overline{R_i} q_i - \varsigma_i \, C_i(q_i, n_i, \delta_i) \right] - g_E \, \theta \geq E_0 \tag{16}$$

$$\sum_{i=1}^{m} \sum_{i'=1}^{m} q_i \Phi_{ii'} q_{i'} + g_v \, \theta \leq V_0 \tag{17}$$

$$\sum_{i=1}^{m} q_i = PROD \tag{18}$$

$$(n_i - 1)\overline{Q_i} \leq q_i \leq n_i \overline{Q_i} \qquad \forall i \in \{1, ..., m\} \tag{19}$$

$$(n_i - 1)\overline{Q_i} + \underline{Q_i} \, \delta_i \leq q_i \qquad \forall i \in \{1, ..., m\} \tag{20}$$

[15] *i.e., it verifies both* $\sum_{i=1}^{m} q_{0i} = PROD$ *and* $\left\{ i \in \{1, ..., m\}, 0 < q_{0i} < \underline{Q_i} \right\} = \varnothing$.

$$q_i \leq \delta_i \overline{Q_i} + (n_i - 1)\overline{Q_i} + \underline{Q_i} \qquad \forall i \in \{1,...,m\} \qquad (21)$$

$$q_i \leq \varsigma_i \, PROD \qquad \forall i \in \{1,...,m\} \qquad (22)$$

$$\underline{Q_i} \, \varsigma_i \leq q_i \qquad \forall i \in \{1,...,m\} \qquad (23)$$

$$\theta \geq 0 \,,\; q_i \geq 0 \,,\; n_i \in \mathbb{N}^* \,,\; \delta_i \in \{0,1\} \,,\; \varsigma_i \in \{0,1\} \;\; \forall i \in \{1,...,m\} \; (24)$$

In this problem, the goal is to find an optimally rebalanced portfolio so as to maximize the value of the non-negative variable θ. Because of the inequalities (16) and (17), θ measures the optimal improvements that can be obtained in terms of increasing returns and decreasing risks in the direction g. Of course, such a rebalanced portfolio must be a feasible one which means that this combination of resource flows $q = (q_1,...,q_i,...,q_m)^T$ and the associated binary and integer variables must satisfy the resource constraint (18) and the technological constraints (19)-(23), i.e. those already used in Problem (P1).

For a country that has to compare several gas monetization policies, this approach provides a simple gauging procedure: applying the same distance function to evaluate the efficiency of the proposed portfolios allows it to rate and compare the performances of the various options. Arguably, the portfolio with the smallest distance possible is deemed the best. If no improvements can be found (i.e., at the optimum, we have $\theta = 0$), then the initial portfolio q_0 belongs to the efficient frontier and is thus reputed efficient. Incidentally, we can remark that this program has a nonempty feasible set.[16]

5.2 Results

In applications, an arbitrary choice must be made for the direction vector g (Briec et al., 2004). In this study, we have chosen the direction $g = (0, E_0)$ which is the "return expansion" approach introduced in Morey and Morey (1999). Accordingly, the thrust is on augmenting the expected amount of perceived resource rents with no increases in the total risk. This methodology has been applied to gauge the efficiencies of the portfolios implemented in these nine countries. In Table 3, we report the obtained results: the optimal improvements and the composition of the optimally rebalanced portfolio.

Several policy recommendations can be derived from these results. First, the results obtained for Bahrain and the UAE confirm the impression derived from the visual observation of Figure 1: the chosen diversification policies exhibit significant inefficiencies. As we are dealing with industrial

---

[16] *The initial portfolio* q_0 *belongs to the feasible set. So,* $\theta = 0$*, and for any* i*:* $q_i = q_{0i}$*,* $n_i = \lceil q_{0i} / \overline{Q_i} \rceil$*,* $\delta_i = 1$ *if* $\underline{Q_i} + (n_i - 1)\overline{Q_i} \leq q_{0i}$*, and* $\varsigma_i = 1$ *if* $q_{0i} > 0$ *satisfy the conditions (18)-(24). Moreover, the expected net present value of that portfolio is* E_0 *and its variance is* V_0 *which is coherent with the satisfaction of the conditions (16) and (17).*

assets, any modification of a previously decided portfolio is likely to generate some rebalancing costs that have not been taken into consideration in this approach. Nevertheless, the magnitude of the gains obtained with the rebalanced portfolios is large enough to suggest that, in both countries, it might be useful to further investigate the possibility of revising the current RBI policies.

Table 3. Efficiency evaluation of the export policy in terms of return expansion

	θ	Allocated shares for the rebalanced portfolio (%)						HHI
		Aluminum smelters	GTL plants	DRI plants	LNG trains	Methanol plants	Urea plants	
Angola	5.1%	-	-	13.1%	58.6%	28.3%	-	44.1%
Bahrain	3275.2%	-	-	-	-	100.0%	-	100.0%
Brunei	1.0%	-	-	-	92.3%	7.7%	-	85.9%
Equatorial Guinea	0.0%	-	-	-	85.4%	14.6%	-	75.1%
Nigeria	9.5%	0.3%	-	-	99.7%	-	-	99.3%
Oman	6.2%	-	-	3.7%	84.5%	11.7%	-	73.0%
Qatar	13.5%	-	-	-	92.3%	7.7%	0.1%	85.7%
Trinidad & Tobago	1.6%	-	-	-	64.6%	35.4%	-	54.2%
U.A.E.	37.9%	4.3%	-	20.9%	56.0%	18.8%	-	39.5%

Note: θ is the achievable improvement defined above. These figures have been obtained using a "return expansion" direction. The allocated shares detail the composition of the portfolio that provides the optimal "return expansion" while preserving the same level of total risk. HHI is the associated Herfindahl-Hirschman Index.

Second, a comparison of the HHI figures listed in Table 1 and Table 3 indicates that diversification is not necessarily a *panacea.* Out of these nine countries, only Angola could derive some benefit from a more diversified use of its gas as its rebalanced portfolio has both a lower HHI score and substantial gains in expected returns. On the contrary, countries like Bahrain, Oman, Nigeria or Qatar could obtain substantial risk-preserving gains in expected returns using significantly less diversified portfolios than those actually implemented. In the case of Bahrain, a complete specialization in methanol processing would even be preferred to the planned portfolio. For Nigeria, an almost complete specialization in LNG could provide a substantial gain without any impact on risk.

Third, we can notice that some countries like Brunei, Equatorial Guinea and Trinidad & Tobago can not expect large gains from a return improving rebalancing of their actual portfolios. In the particular case of Equatorial Guinea, no improvement can be obtained, meaning that this country's portfolio belongs to the efficient frontier. This latter finding may be explained by the fact that, in this country, the decision to construct both an LNG train and a methanol plant resulted from an integrated planning approach. For Brunei, the improved portfolio solely involves a minor rebalancing between the shares of the chosen technologies: methanol and LNG. Concerning Trinidad & Tobago, these findings can be used to inform a local debate about the opportunity to install an aluminum smelter. During the last decade, this large project generated a controversial debate in the Caribbean nation before being officially canceled by a governmental decision in 2010. Interestingly, the government's motivation for halting this project explicitly mentioned concerns about the optimal use of the nation's gas resources. Our results indicate that aluminum smelting is not part of the country's optimal portfolio and thus provide some support for that decision.

Lastly, the relative attractiveness of the various technologies deserves a comment. We focus on the GTL technology because this extremely capital intensive technology is experiencing an upsurge in interest. In addition to the large GTL plants recently installed in Nigeria and Qatar, several GTL projects are currently under review in Algeria, Bolivia, Egypt, Turkmenistan and Uzbekistan (MHEB, 2008; IEA, 2010, p. 145). Interestingly, our findings indicate that the GTL option is never selected in any of the optimally rebalanced portfolios listed in Table 3. Moreover, a meticulous examination of the composition of the portfolios located on the nine efficient frontiers has been carried out and has confirmed that the GTL technology is never chosen in these efficient portfolios. The fact that the export revenues derived from this technology are highly correlated with those of LNG though being far less lucrative can explain these poor results. These findings may have a country-specific nature. Nevertheless, they suggest that it might be preferable to initiate some further studies aimed at meticulously assessing the economics of these GTL projects before authorizing their constructions.

6. Concluding remarks

For small open economies which are unusually well-endowed with natural resources, the positive role played by export diversification in improving economic outcomes is part of the conventional wisdom among analysts, policy makers and the population at large. In this paper, we analyze the economics of an export diversification strategy centered on the deployment of resource-based industries. In this regard, attention is focused on the extent to which a wisely selected RBI strategy may reduce the variability of the country's export earnings and/or enhance the expected level of perceived resource rents.

The challenge of this paper is to specify an adapted MVP model that explicitly takes into consideration the main features of resource-based industries (differences in the processing costs, existence of economies of scale at the plant level, and lumpiness). We believe that this model-based approach is able to provide valuable guidance for the decision makers involved in the design of an export-oriented RBI strategy, especially if the industries under scrutiny are aimed at being regrouped within an Export Processing Zone (EPZ), a rather popular policy instrument that usually involves generous and long-term tax holidays and concessions to the firms.

In a case study focusing on natural gas, the paper analyzes the optimal export portfolios that can be considered by a country endowed with significant deposits of natural gas. From a methodological perspective, this study allows us to present some clarifications on the practical implementation of the proposed approach (e.g. modeling of the random export revenues). At an empirical level, we have evaluated the efficient export frontier of nine gas-rich economies. We observe that the countries' efficient frontier varies with the countries' endowment and that a larger endowment offers many more options for policy planners. Besides which, our findings confirm that the raw exports of natural gas based on cryogenic facilities (i.e., LNG) systematically provide the highest returns, but also the highest risk. In addition, we conduct a quantitative assessment of the efficiency of the export portfolio implemented in these nine countries. The results indicate that, in all countries but one (Equatorial

Guinea), the RBI strategy that has been implemented is outperformed by an optimally rebalanced portfolio. Although this application focuses on the case of natural gas, it should be clear that a similar approach could apply to other resources as well (for example: oil and petrochemicals, cotton and textile, agricultural commodities and the agro-industries...).

As in any modeling effort, we made some simplifying assumptions. The two main ones are that volume variability can be neglected compared to those of international prices and that the flow of resource is determined exogenously without taking into consideration that sector's economics (e.g., depletion in case of a non-renewable resource). It is clearly of interest to relax both assumptions in future research.

References

Ai, C., Chatrath, A., Song, F., 2006. On the Comovement of Commodity Prices. *American Journal of Agricultural Economics*, 88(3), 574–588.

Aizenman, J., Marion N., 1991. Policy uncertainty, persistence and growth. *NBER Working Paper 3848*, National Bureau of Economic Research.

Alexeev, M., Conrad, R., 2009. The Elusive Curse of Oil. *The Review of Economics and Statistics*, 91(3), 586–598.

Alwang, J., Siegel, P.B., 1994. Portfolio Models and Planning for Export Diversification: Malawi, Tanzania and Zimbabwe. *Journal of Development Studies*, 30(2), 405–422.

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23(2), 193–212.

Asche, F., Osmundsen, P., Tveterås, R., 2002. European market integration for gas? Volume flexibility and political risk. *Energy Economics*, 24(3), 249–265.

Auty R.M., 1988a. Oil-exporters' disappointing diversification into resource-based industry. The external causes. *Energy Policy*, 16(3), 230–242.

Auty R.M., 1988b. State enterprise and resource based industry in oil exporting countries. *Resources Policy*, 14(4), 275–287.

Auty, R.M., 1990. *Resource-Based Industrialization: Sowing the Oil in Eight Developing Countries*, Oxford University Press, Oxford.

Auty, R.M., 2001. *Resource Abundance and Economic Development*, Oxford University Press, Oxford.

Auty R.M., Gelb, A., 1986. Oil Windfalls in a Small Parliamentary Democracy: Their Impact on Trinidad and Tobago. *World Development*, 14(9), 1161–1175.

Bernanke, B.S., 1983. Irreversibility, uncertainty and cyclical investment. *The Quarterly Journal of Economics*, 98(1), 85–106.

Berndt, E.R., Hall, B.H., Hall, R.E., Hausmann, J.A., 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3/4, 653–665.

Bertinelli, L., Heinen A., Strobl, E., 2009. Export Diversification and Price Uncertainty in Developing Countries: A Portfolio Theory Approach. Working Paper available at SSRN: http://ssrn.com/abstract=1327928.

Blattman, C., Hwang, J., Williamson, J.G., 2007. Winners and losers in the commodity lottery: The impact of terms of trade growth and volatility in the Periphery 1870–1939. *Journal of Development Economics*, 82(1), 156–179.

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalised ARCH model. *The Review of Economics and Statistics*, 72(3), 498–505.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A Capital Asset Pricing Model with Time-Varying Covariances. *The Journal of Political Economy*, 96(1), 116–131.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi maximum likelihood estimation of dynamic models with time varying covariance. *Econometric Reviews*, 11, 143–172.

BP, 2011. *BP Statistical Review of World Energy*, BP, London.

Brainard, W.C., Cooper, R.N., 1968. Uncertainty and Diversification in International Trade. *Food Research Institute Studies in Agricultural Economics: Trade and Development*. 8: 257-285.

Briec, W., Kerstens, K., Lesourd, J.-B, 2004. Single period Markowitz portfolio selection, performance gauging and duality: A variation on the Luenberger shortage function. *Journal of Optimization Theory and Applications*, 120(4), 1–27.

Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1996. A test for independence based on the correlation dimension. *Econometric Reviews*, 15, 197–235.

Brock, W.A., Hsieh, D.A., LeBaron, B., 1993. *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. MIT Press, Cambridge (MA).

Brüggemann, R., Lütkepohl, H., 2001. Lag selection in subset VAR models with an application to a US monetary system, in: Friedmann, R., Knüppel, L., Lütkepohl, H., (Eds.), *Econometric Studies: A Festschrift in Honour of Joachim Frohn*. LIT, Münster, pp. 107–128.

Bunn, D.W., 2004. *Modelling Prices in Competitive Electricity Markets*, Wiley, London.

Caceres, L.R., 1979. Economic Integration and Export Instability in Central America: A Portfolio Model. *Journal of Development Studies*, 15(3), 141–153.

Corden, W.M., Neary, J.P., 1982. Booming Sector and De-Industrialisation in a Small Open Economy. *The Economic Journal,* 92(368), 825–848.

Deb, P., Trivedi, P.K., Varangis, P., 1996. The Excess Comovement of Commodity Prices Reconsidered. *Journal of Applied Econometrics*, 11, 275–91.

Davis, J., Ossowski, R., Daniel, J., Barnett, S., 2001. Stabilization and Savings Funds for Nonrenewable Resources Experience and Fiscal Policy Implications. Occasional Paper 205. International Monetary Fund, Washington, DC.

Devlin, J., Lewin, M., 2005. Managing Oil Booms and Busts in Developing Countries, in: Aizenman, J., Pinto, B. (Eds.), *Managing economic volatility and crises: a practitioner's guide*. Cambridge University Press, Cambridge, pp. 186–209.

Doornik, J.A., Hansen, H., 2008. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939

Engle, R.F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987–1007.

Engle, R.F., 2002. Dynamic Conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20, 339–350.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11, 122–150.

Engle, R.F., Sheppard, K., 2001. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. *NBER Working Paper 8554*, National Bureau of Economic Research.

ESMAP, 1997. *Commercialization of marginal gas fields*. Report no. 201/97. Joint UNDP/World Bank Energy Sector Management Assistance Programme. The World Bank. Washington, D.C.

ESMAP, 2004. *Strategic Gas Plan for Nigeria*. Report, February 2004. Joint UNDP/World Bank Energy Sector Management Assistance Programme. The World Bank. Washington, D.C.

Frankel, J.A., 2010. The Natural Resource Curse: A Survey. *NBER Working Paper 15836*, National Bureau of Economic Research.

Gelb, A., 1988. *Oil Windfalls: Blessing or Curse*, Oxford University Press for the World Bank, New York.

Geman, H., 2005. *Commodities and Commodity Derivatives*, Wiley, London.

Gonzalo, J., 1994. Comparison of five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, 60(1-2), 203–233.

Hausmann, R., Rigobon, R., 2003. An Alternative Interpretation of the "Resource Curse": Theory and Policy Implications, in: Davis, J.M, Ossowski, R., Fedelino, A. (Eds.), *Fiscal Policy Formulation and Implementation in Oil-Producing Countries,* IMF Press, Washington, D.C.

Hassan, M. F., 1975. *Economic growth and employment problems in Venezuela: An analysis of an oil-based economy*, Praeger, New-York.

Hirschman, A.O., 1958. *The strategy of economic development*, Yale University Press, New Haven CT.

Horst, R., Tuy, H., 1996. *Global optimization: Deterministic Approaches*, third ed., Springer-Verlag, Berlin.

Hsieh, D.A., 1991. Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance*, 46, 1839–1877.

Humphreys, M., Sandbu, M.E., 2007. The Political Economy of Natural Resource Funds, in: Humphreys, M., Sachs, J., Stiglitz, J. (Eds.), *Escaping the Resource Curse*. Columbia University Press, New York, , pp. 194–234.

IEA (International Energy Agency), 2010. *World Energy Outlook 2010*. OECD/IEA, Paris.

Isham, J., Pritchett, L., Woolcock, M., Busby, G., 2005. The Varieties of Resource Experience: Natural Resource Export Structures and the Political Economy of Economic Growth. *World Bank Economic Review*, 19(2), 141–174.

Johansen, S., 1988. Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, 12(2-3), 231–254.

Johansen, S., 1995. *Likelihood-Based Inference in Cointegrated Vector Autogressive Models*, Oxford University Press, Oxford.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52(2), 169–210.

Klein, B., Crawford, R.G., Alchian, A.A., 1978. Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. *Journal of Law and Economics,* 21(2), 297–326.

Labys, W.C., Lord, M.J., 1990. Portfolio Optimisation and the Design of Latin American Export Diversification Policies. *Journal of Development Studies*, 26(2), 260–277.

Leybourne, S.Y., Lloyd, T.A., Reed G.V., 1994. The excess comovement of commodity prices revisited. *World Development*, 22(11), 1747–1758.

Love, J., 1979. A model of trade diversification based on the Markowitz model of portfolio analysis. *Journal of Development Studies*. 15(2), 233–241.

Love, J., 1983. Concentration, diversification and earnings instability: Some evidence on developing countries' exports of manufactures and primary products. *World Development*, 11(9), 787–793.

Love, J., 1986. Commodity concentration and export earnings instability: A shift from cross-section to time series analysis. *Journal of Development Economics*. 24(2), 239–248.

Lütkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

MacKinnon, J.G., 1996. Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11, 601–618.

Markowitz, H.M., 1952. Portfolio selection. *The Journal of Finance*, 7(1), 77–91.

Markowitz, H.M., 1959. *Portfolio selection: efficient diversification of investments*, Wiley, New York.

Mendoza, E.G., 1997. Terms-of-Trade Uncertainty and Economic Growth. *Journal of Development Economics*, 54(2), 323–356.

MHEB, 2008. *Estrategia Boliviana de Hidrocarburos*. Ministerio de Hidrocarburos y Energía de Bolivia, La Paz – Bolivia.

Morey, M.R., Morey, R.C., 1999. Mutual Fund Performance Appraisals: A Multi-Horizon Perspective with Endogenous Benchmarking, *Omega*, 27, 241–258.

Morrison, D.F., 1967. *Multivariate Statistical Methods*, McGraw-Hill Book Company, New York.

Owens, T., Wood A., 1997. Export-oriented industrialization through primary processing? *World Development*, 25(9), 1453–1470.

Pearson, S.R., 1970. *Petroleum and the Nigerian economy*, Stanford University Press, Stanford CA.

Perold, A.F., 1984. Large-Scale Portfolio Optimization. *Management Science*, 30(10), 1143–1160.

Pindyck, R.S., 1991. Irreversibility, uncertainty, and investment. *Journal of Economic Literature*, 129(3), 1110–1148.

Pindyck, R.S., Rotemberg, J., 1990. The excess co-movement of commodity prices. *The Economic Journal*, 100(403), 1173–1189.

Prebisch, R., 1950. *The Economic Development of Latin America and Its Principal Problems*, United Nations, New York.

Roemer, M., 1979. Resource-based industrialization in the developing countries : A survey. *Journal of Development Economics*, 6(2), 163–202.

Roques, F.A., Newbery, D.M., Nuttall, W.J., 2008. Fuel mix diversification incentives in liberalized electricity markets: a mean-variance portfolio theory approach. *Energy Economics,* 30(4), 1831–1849.

Rosenstein-Rodan, P. 1943. Problems of industrialisation of Eastern and South-Eastern Europe. *The Economic Journal*, 53(210/211), 202–211.

Ross, M.L., 1999. The Political Economy of the Resource Curse. *World Politics*, 51(2), 297–322.

Sachs, J.D., Warner, A.M., 1995. revised 1997, 1999. Natural resource abundance and economic growth. *NBER Working Paper 5398*, National Bureau of Economic Research, Cambridge, MA.

Sahinidis, N.V., 1996. BARON: A general purpose global optimization software package. *Journal of Global Optimization,* 8(2), 201–205.

Singer, H.W, 1950. US Foreign Investment in Underdeveloped Areas: The Distribution of Gains between Investing and Borrowing Countries. *American Economic Review Papers and Proceedings*, 40(2), 473–485.

Stevens, P., 2003. Resource Impact: curse or blessing? A literature survey. *Journal of Energy Literature*, 9(1), 3–42.

Tawarmalani, M., Sahinidis, N.V., 2004. Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Mathematical Programming*, 99(3), 563–591.

van der Ploeg, F., 2011. Natural Resources: Curse or Blessing? *Journal of Economic Literature*, 49(2), 366–420.

van der Ploeg, F., Poelhekke, S., 2009. Volatility and the Natural Resource Curse. *Oxford Economic Papers*, 61(4), 727–760.

Xue, H., Xu, C., Feng, Z., 2006. Mean–variance portfolio optimal problem under concave transaction cost. *Applied Mathematics and Computation*, 174(1), 1–12.

Appendix A

Proof of Lemma 1

We have to prove that there exists at least one possible industrial configuration that jointly satisfies the equations (3), (4) *and* (5). *The case of a full specialization of the technology with the smallest implementable size provides an interesting candidate. Let* $k = \arg\text{Min}_i \left\{\underline{Q_i}\right\}$, *we can prove that a multi-plant industrial configuration with: (i)* $n_i = 1$ *and* $q_{i1} = 0$ *for all the other goods* $i \in \{1,...m\} \setminus \{k\}$, *and (ii)* n_k *plants of type* k *with sizes defined hereafter, satisfies all the conditions. Concerning the sizes of the plants producing good* k, *two cases must be distinguished. Let* $n_k = \left\lceil PROD / \overline{Q_k} \right\rceil$ *where* $\lceil . \rceil$ *is the ceiling function, and* $r_k = PROD - (n_k - 1)\overline{Q_k}$. *If* $r_k \geq \underline{Q_k}$ *(Case 1), then a multi-plant industrial configuration with* n_k *plants of type* k *of which* $(n_k - 1)$ *plants of maximum size* $\overline{Q_k}$ *and a last plant of size* r_k *satisfies all the conditions. Otherwise (Case 2), we have* $r_k < \underline{Q_k}$ *and an industrial configuration with* n_k *plants of type* k *of which* $n_k - 2$ *plants of size* $\overline{Q_k}$, *one plant of size* $\underline{Q_k}$ *and one plant of size* $r_k + \left(\overline{Q_k} - \underline{Q_k}\right)$ *satisfies all the conditions. In the latter case,* $PROD \geq \underline{Q_k}$ *insures that* $n_k \geq 2$ *when* $r_k < \underline{Q_k}$. *In addition, the range of implementable sizes is assumed to be large enough to verify* $2\underline{Q_k} \leq \overline{Q_k}$ *which, together with the fact that* $r_k \geq 0$, *insures that a plant of size* $r_k + \left(\overline{Q_k} - \underline{Q_k}\right)$ *is larger than the minimum implementable size* $\underline{Q_k}$. *As we have* $r_k < \underline{Q_k}$, *the capacity* $r_k + \left(\overline{Q_k} - \underline{Q_k}\right)$ *is also smaller than the maximum implementable size* $\overline{Q_k}$.

Q.E.D.

Proof of Proposition 1

The proof is based on three successive steps.

<u>STEP #1:</u> To begin with, we provide a characterization for the cost-minimizing allocation of an exogenously determined flow of resource S_i *with* $2\underline{Q_i} \leq S_i \leq 2\underline{Q_i}$ *that involves exactly two plants. We consider a given pair of plants* $j \in \{1, 2\}$, *each processing a strictly positive flow* $\underline{Q_i} \leq q_{ij} \leq \underline{Q_i}$ *at a cost* $c_i(q_{ij})$. *To avoid index permutations, we assume that plants are ordered in decreasing sizes. So, we are facing the following non-convex non-linear optimization problem (NLP):*

$$\min \quad c_i(q_{i1}) + c_i(q_{i2}) \tag{A.1}$$

$$\text{s.t.} \quad q_{i1} + q_{i2} = S_i \tag{A.2}$$

$$q_{i1} \geq q_{i2} \qquad \text{(A.3)}$$

$$q_{ij} \in \left[\underline{Q_i}, \overline{Q_i}\right] \qquad \forall j \in \{1,2\} \qquad \text{(A.4)}$$

Using (A.2)*, we can reformulate this NLP as a single-variable optimization problem and let* $\alpha = q_{i1} / S_i$ *be that variable. Equation* (A.2) *imposes that* $q_{i2} = (1-\alpha) S_i$. *Because of* (A.3)*,* α *must verify* $\alpha \geq 1/2$. *Because of* (A.4)*, we have* $\alpha \in \left[\underline{Q_i}/S_i, \overline{Q_i}/S_i\right]$ *and* $\alpha \in \left[1-\overline{Q_i}/S_i, 1-\underline{Q_i}/S_i\right]$. *Given that* $S_i \leq 2\overline{Q_i}$, *we have* $1-\overline{Q_i}/S_i \leq 1/2$. *Moreover, we have* $\underline{Q_i}/S_i < 1/2$ *because* $2\underline{Q_i} \leq S_i$. *Accordingly, the NLP can be simplified as follows: find* $\alpha \in \left[1/2, Min\left\{\overline{Q_i}/S_i, 1-\underline{Q_i}/S_i\right\}\right]$ *that minimizes the overall cost* $c_i(\alpha q_i) + c_i((1-\alpha) q_i)$. *Given that* $S_i \leq 2\overline{Q_i}$ *and that* $2\underline{Q_i} \leq S_i$, *this latter interval is nonempty. Given that* c_i *is a twice continuously differentiable function, we can evaluate the derivative of* $c_i(\alpha q_i) + c_i((1-\alpha) q_i)$ *with respect to* α. *Because of the concavity of* c_i, *we have* $c_i^{'}(\alpha q_i) \leq c_i^{'}((1-\alpha) q_i)$ *indicating that the total cost function is strictly decreasing for any* $\alpha \geq 1/2$. *Hence, the optimal solution* α^* *is given by the upper bound i.e.,* $\alpha^* = Min\left\{\overline{Q_i}/S_i, 1-\underline{Q_i}/S_i\right\}$. *Using words, this result indicates that: (i) if the quantity to be processed is large enough (i.e.* $\overline{Q_i} + \underline{Q_i} \leq S_i$*), we have* $\alpha^* = \overline{Q_i}/S_i$ *indicating that the plant* $j = 1$ *has the maximum implementable size; (ii) otherwise (i.e.* $\overline{Q_i} + \underline{Q_i} > S_i$*), we have* $\alpha^* = 1 - \underline{Q_i}/S_i$ *indicating that the plant* $j = 2$ *has the minimum implementable size.*

As a corollary, this result provides an interesting characterization: for any S_i *with* $2\underline{Q_i} \leq S_i \leq 2\overline{Q_i}$, *the cost-minimizing allocation of* S_i *among two plants imposes that at least one plant has a size equal to the bounds (either* $\underline{Q_i}$ *or* $\overline{Q_i}$ *).*

<u>STEP #2:</u> Now, we consider the number of processing plants n_i *as a parameter. Assume a given flow of resource to be processed* q_i *with* $q_i \geq \underline{Q_i}$ *using* n_i *plants* $j \in \{1, \ldots, n_i\}$ *that processes strictly positive flows* q_{ij} *with* $\underline{Q_i} \leq q_{ij} \leq \overline{Q_i}$. *The plant's cost is* $c_i(q_{ij})$. *We also assume that* q_i *and* n_i *jointly verify* $n_i \underline{Q_i} \leq q_i \leq n_i \overline{Q_i}$. *If we assume that a cost-minimizing allocation* $(q_{ij})_{j \in \{1,\ldots,n_i\}}$ *of the overall flow* q_i *among these* n_i *plants has at least two plants indexed* k_1 *and* k_2 *with* $\underline{Q_i} < q_{ik_1} < \overline{Q_i}$ *and* $\underline{Q_i} < q_{ik_2} < \overline{Q_i}$, *then we have a contradiction with the fact that* $(q_{ij})_{j \in \{1,\ldots,n_i\}}$ *is a cost-minimizing*

allocation (because applying the characterization obtained in Step #1 to the plants k_1 and k_2 indicates that, ceteris paribus, it is possible to find an allocation with a lower cost capable to process that good). Accordingly, any cost-minimizing allocation of the overall flow q_i among these n_i plants has at most one unique plant $k \in \{1,...,n_i\}$ processing q_{ik} with $\underline{Q_i} < q_{ik} < \overline{Q_i}$.

<u>STEP #3:</u> For a given level of resource to be processed q_i with $q_i \geq \underline{Q_i}$, possible values for the integer n_i have to be chosen in a bounded set: $N := \left\{ n_i \in \mathbb{N} : n_i \underline{Q_i} \leq q_i \leq n_i \overline{Q_i} \right\}$. We are now going to compare the cost-minimizing allocations $\left(q_{ij}^*\right)_{j \in \{1,...,n_i\}}$ obtained with these various n_i. Denote $n_i^* = \underset{n_i \in N}{ArgMin} \left\{ \sum_{j=1}^{n_i} c_i\left(q_{ij}^*\right) \right\}$ the number of plants that provides the (or one of the) least costly configuration $\left(q_{ij}^*\right)_{j \in \{1,...,n_i^*\}}$ among all the possible integers $n_i \in N$ and their associated cost-minimizing configurations $\left(q_{ij}^*\right)_{j \in \{1,...,n_i\}}$. Hereafter, we provide a characterization for the optimal (cost-minimizing) configuration $\left(q_{ij}^*\right)_{j \in \{1,...,n_i^*\}}$.

We note that c_i is a single-variable concave cost function with $c_i(0) = 0$. Thus, c_i is subadditive. Given that $2\underline{Q_i} \leq \overline{Q_i}$, we have $2c_i\left(\underline{Q_i}\right) \geq c_i\left(2\underline{Q_i}\right)$ indicating that a single plant design is always preferable (from a cost-minimizing perspective) to process $2\underline{Q_i}$. Accordingly, any integer $n_i \in N$ with a cost-minimizing configuration $\left(q_{ij}^*\right)_{j \in \{1,...,n_i\}}$ including more than two plants of minimum size $\underline{Q_i}$ cannot be optimal (because these two plants of size $\underline{Q_i}$ could be concatenated within a single plant indicating that there exists at least one configuration involving $n_i - 1$ plants that is capable to process q_i at a lower cost).

So, this last result (together with those obtained in Step #2) indicates that the optimal number of plants $n_i^* \in N_i$ must satisfy at least one of these four conditions:

case 1: $n_i^* \overline{Q_i} = q_i$

case 2: $\left(n_i^* - 1\right)\overline{Q_i} + \underline{Q_i} = q_i$

case 3: $\left(n_i^* - 1\right)\overline{Q_i} + r_i = q_i$ *with* $\underline{Q_i} < r_i < \overline{Q_i}$

case 4: $\left(n_i^* - 2\right)\overline{Q_i} + r_i + \underline{Q_i} = q_i$ *with* $\underline{Q_i} < r_i < \overline{Q_i}$

For any level of resource to be processed q_i *with* $q_i \geq \underline{Q_i}$*, the configuration listed in Proposition 1 satisfies one of the conditions and is thus a least cost organization.* *Q.E.D.*

Proof of Proposition 2:

When considering the integer and binary variables (n_i), (δ_i) *and* (ς_i) *as parameters, this problem turns into a nonlinear optimization problem (NLP) that has an interesting form:*

$$\text{Problem } (NLP_{n,\delta,\varsigma}) \quad \max \quad \overline{R}^T q - Cost_{n,\delta,\varsigma}(q) - \frac{\lambda}{2} q^T \Phi q \tag{A.5}$$

$$\text{s.t.} \quad q \in D_{n,\delta,\varsigma} \cap S_n \tag{A.6}$$

where $Cost_{n,\delta,\varsigma}(q) = \sum_{i=1}^{m} \varsigma_i \, C_i(q_i, n_i, \delta_i)$ *is the sum of twice continuously differentiable univariate concave functions. The set* $D_{n,\delta,\varsigma}$ *is a polytope defined by a series of linear inequalities associated with the collection of linear constraints of type (8), (10), (11), (12), (13) and (14). The set* $S_n = \left\{ (n_i - 1)\overline{Q_i} \leq q_i \leq n_i \overline{Q_i}, \; i = 1,...,m \right\}$ *is a rectangle of upper and lower bounds on the vector* q *that corresponds to the constraints of type (9).*

If the feasible set $D_{n,\delta,\varsigma} \cap S_n$ *is nonempty, the objective function is continuous and real valued on a closed and bounded set and thus the problem* $NLP_{n,\delta,\varsigma}$ *has a solution (Weierstrass Theorem). From a computational perspective, the variance-covariance matrix is positive semi definite and the plants' cost functions are concave. So, the problem* $NLP_{n,\delta,\varsigma}$ *is a "well behaved" nonlinear, non convex, optimization problem: a special case of Difference of Convex (DC) programming as defined in Horst and Tuy (1996).*[17]

In addition, the number of combinations of integer and binary variables that have to be considered is thus bounded because any integer value n_i *larger than* $\left(\left\lceil PROD / \overline{Q_i} \right\rceil + 1 \right)$ *cannot jointly satisfies equations (8) and (9).*

Moreover, we can prove that there exists at least one combination of discrete parameters that verifies the conditions for a nonempty feasible set.. As $PROD \geq \text{Min}_i \left\{ \underline{Q_i} \right\}$*, Lemma 1 (cf. proof) suggests a candidate: a full specialization in the good* $k = \arg\text{Min}_i \left\{ \underline{Q_i} \right\}$*. If we consider the discrete parameters:* $n_i = 1$, $\delta_i = 0$, *and* $\varsigma_i = 0$ *for any* $i \in \{1,...m\} \setminus \{k\}$ *together with* $n_k = \left\lceil PROD / \overline{Q_k} \right\rceil$

[17] *Dedicated algorithms have recently been proposed to solve such a DC problem. For example, Xue et al. (2006) have constructed a branch-and-bound scheme using linear underestimating functions of the univariate concave cost functions aimed at creating an outer underestimate relaxation of the original problem.*

plants of type k, $\varsigma_i = 1$ *and* $\delta_k = min\left(\left\lfloor \frac{PROD - (n_i - 1)\overline{Q_k}}{\underline{Q_k}} \right\rfloor, 1\right)$, *then the vector* q *with* $q_i = 0$ *for any* $i \in \{1,...m\} \setminus \{k\}$ *and* $q_k = PROD$ *verifies all the conditions (8), (9) (10), (11), (12), (13) and (14). So, for these discrete parameters, the feasible* $D_{n,\delta,\varsigma} \cap S_n$ *is nonempty.*

So, given that (i) the number of combinations that are worth being considered is bounded, and (ii) there exists at least one combination of discrete parameters that provides a real valued solution, an enumeration of the solutions of $\left(\mathrm{NLP}_{n,\delta,\varsigma}\right)$ *for the various combinations of discrete parameters provides the global solution to the problem (P1).* *Q.E.D.*

Appendix B

In this Appendix, we detail the data used in our numerical analyses.

Table B-1. Cost parameters for the individual gas processing plants

Gas use (gauging equipment)	Gas input (Mcf/ton)	Range of implementable processing capacities (ktpa - 10^3 tons per annum) Minimum	Maximum	Investment cost function $C_i(q_{ij}) = \alpha_i . q_{ij}^{\beta_i}$ (US$ with q_{ij} in tpa) α_i	β_i	O&M cost (US$/ton)	Freight (US$/ton)	Cost of raw minerals (if any) (% of output price)
Aluminum *(line pot)*	91.13	50.00	386.00	12 424.33	0.941	745.25	28.57	26.74
Gas-to-Liquid *(Fischer-Tropsch reactor)*	71.82	110.99	838.61	3 517.74	1.000	44.40	22.20	-
Direct Reduced Iron *(shaft furnace)*	12.17	310.00	1 950.00	2 276.56	0.840	16.16	17.14	63.00
Liquefied Natural Gas *(liquefaction train)*	55.35	2 500.00	7 100.00	3 843.43	0.853	9.71	29.00	-
Methanol *(methanol reactor)*	31.76	204.00	3 400.00	3 023.30	0.875	41.00	24.00	-
Urea *(urea reactor)*	21.61	170.00	1 500.00	4 161.45	0.832	62.00	22.86	-

Note #1: All cost figures are in 2010 US dollars. All plants are assumed to be at a port location with adequate infrastructure. For aluminum, the cost figures correspond to an integrated project (smelter + gas power plant). For mineral-related activities (Aluminum and DRI), an assumption has been made on the price of the raw mineral to be processed: alumina price in $/t is assumed to be equal to 14% of those of aluminum (Rio Tinto) and, it is assumed that 1.91 t of alumina are required for each t of aluminum (US DoE). Concerning DRI, we assume that 1.5 ton of fine iron ore is required for each ton of DRI (ESMAP, 1997). The price of iron ore in $/t is assumed to be equal to 42% of those of scrap steel (the mean value observed during the last five years). For GTL, a conversion factor of 1 barrel of diesel oil per day = 49.33 metric tons per year has been used. For LNG, prices and processing costs are frequently given in US$ per MMBTU and the following conversion has been used: 1 ton of LNG = 48.572 MMBTU.

Note #2: These data have been gathered from institutions (The Energy Technology Systems Analysis Program of the International Energy Agency, The Energy Sector Management Assistance Program, The U.S. Department of Energy), associations (Cedigaz, International Aluminum Institute, GIIGNL, Society of Petroleum Engineers) and companies (Qatar Fertilizer Co., HYL/Energiron, Marathon, Midrex, Rio Tinto, Sasol, Shell, Stamicarbon). The inter-industry coherence has been checked using proprietary detailed cost engineering studies available at IFP Energies Nouvelles, a large French R&D center entirely focused on the energy industries.

Appendix C - An empirical model of future export revenues

This section details the construction of a data-driven time series model of the data generating process associated with the international prices of six commodities. We proceed as follows. First, a concise description of the data set is given. Second, an appropriate methodology is proposed. Lastly, the proposed specification is estimated and results are commented.

C.1 Data, descriptive statistics, unit root tests and cointegration analysis

a - A Preliminary Look at the Data

The data employed in this study consists of monthly prices of six commodities: aluminum (hereafter named ALU), diesel oil (DIES), Direct Reduced Iron (DRI), natural gas in the European Union (GAS), methanol (MET), and urea (UREA). These prices have been collected from January 1990 to February 2010. Data has been gathered from: the commodity price data published by the World Bank (GAS, UREA), the IMF Primary Commodity Prices (ALU), Platt's quotations for methanol (MET) and diesel oil (DIES).[18] The price series for DRI is derived from the World Bank and the US Geological Survey.[19] All these prices have been transformed into 2010 US dollars (reference January 2010) using the Consumer Price Index published by the U.S. Bureau of Labor Statistics. Figure C-1 provides plots of these monthly prices and Table C-1 summarizes the descriptive statistics of these series. The coefficient of variation, that measures the degree of variation relative to the mean price, ranges from 20.1% (aluminum) to 53.6% (diesel oil). On this basis, mineral commodities (aluminum and DRI) seem less variable than fuels (natural gas and diesel oil) and gas-based chemicals (methanol and urea). With the exception of aluminum, the kurtosis exceeds three, which suggests that leptokurtic distributions may be indicated.

Table C-1. Summary statistics (price levels)

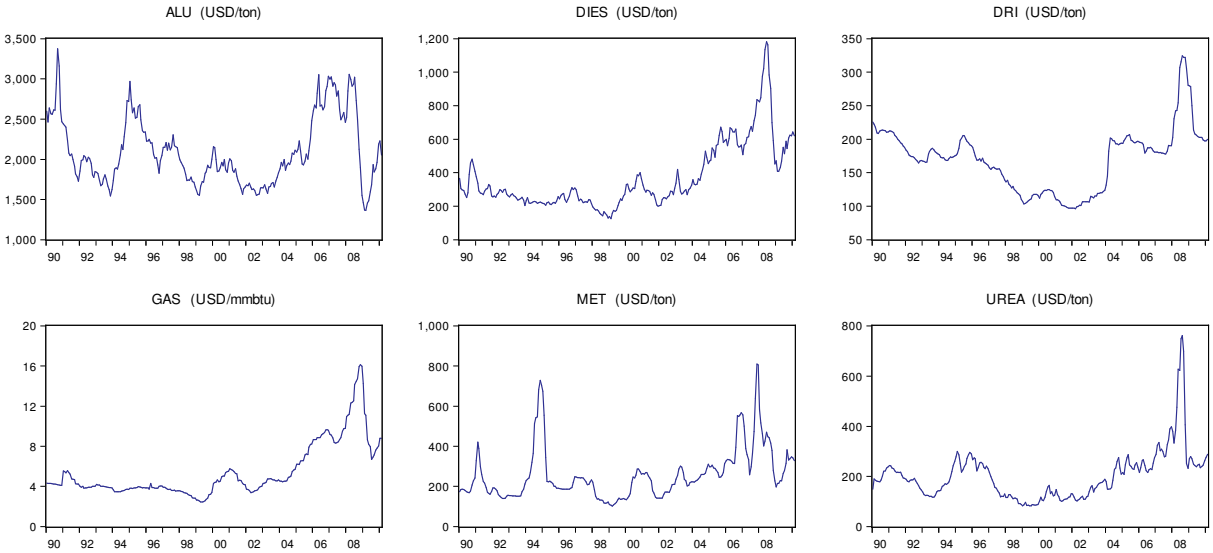
	ALU	DIES	DRI	GAS	MET	UREA
Observations	242	242	242	242	242	242
Mean	2095.469	368.342	169.540	5.456	259.601	202.758
Median	1993.040	291.302	176.079	4.289	227.120	186.924
Maximum	3375.223	1181.718	324.630	16.127	810.242	761.568
Minimum	1363.104	124.780	96.310	2.438	101.777	82.464
Standard deviation	421.408	197.268	46.526	2.738	130.107	101.920
Coefficient of variation	0.201	0.536	0.274	0.502	0.501	0.503
Skewness	0.753*	1.666***	0.566	1.803***	1.828***	2.472***
Excess Kurtosis	-0,277	2.938*	0.925	3.070**	3.674**	9.802***

Note: Asterisks indicate rejection of the null of a coefficient equal to zero at 0.10*, 0.05** and 0.01*** levels, respectively.

[18] *The quotations used are Methanol Spot Rotterdam, and Diesel Oil n°2 New York Cargo Spot, respectively.*

[19] *DRI can be reduced to steel in electric arc furnaces with varying inputs of scrap steel. Because of this flexibility, DRI prices are reputed to be very close to those of scrap steel. Because of the lack of publicly available price series for DRI, a proxy has been constructed by multiplying: (1) the World Bank's steel product price index (with reference 100 in year 2000), and (2) the price of scrap steel in 2000 as published by the US Geological Survey: 96 $/ton.*

Figure C-1. Data plots



Note: All prices are in 2010 USD.

From Figure C-1, a visual inspection suggests that the commodity price series may exhibit some co-movements. In addition, the correlation matrix (see Table C-2) indicates the presence of positive correlations among these prices. All of these correlation coefficients significantly differ from zero. A likelihood ratio test of the null hypothesis that the correlation matrix is equal to the identity matrix has also been conducted and allowed us to reject that hypothesis. Hence, these correlations are, as a group, statistically significant. From an economic perspective, these positive figures suggest that an export-oriented diversification through RBI cannot totally eradicate the export earnings variability. Nonetheless, we can seek to mitigate its amplitude.

Table C-2. Unconditional correlations (price levels)

	ALU	DIES	DRI	GAS	MET	UREA
ALU	1.000					
DIES	0.526***	1.000				
DRI	0.491***	0.611***	1.000			
GAS	0.380***	0.863***	0.677***	1.000		
MET	0.577***	0.563***	0.368***	0.520***	1.000	
UREA	0.596***	0.798***	0.768***	0.735***	0.575***	1.000
$\chi^2(15)$ = 1102.002 *(0.000)*						

Note: *** indicates significance at the 0.01 level. $\chi^2(15)$ is the Bartlett statistic associated with a likelihood ratio test of the null hypothesis that the correlation matrix is equal to the identity matrix described in Morrison (1967, p. 113). The associated *p*-value is in brackets.

b - Unit root tests

As is now standard practice in the analysis of commodity price time series, this subsection begins with the examination of unit roots of all the variables. The presence of unit root indicates that a time series is nonstationary. Testing for the order of integration in individual series is based on the traditional

Augmented Dickey Fuller (ADF) test and the Philips-Perron (PP) test. In these tests, the existence of a unit root is the null hypothesis.

Table C-3. Unit root tests

		ADF			PP		
		Constant & Trend	Constant	None	Constant & Trend	Constant	None
ALU	Level	-2.955	-2.951	-0.741	-2.871	-2.862	-0.790
	First difference	-11.616***	-11.643***	-11.665***	-11.616***	-11.643***	-11.665***
DIES	Level	-3.530	-2.408	-0.814	-2.980	-1.856	-0.590
	First difference	-7.487***	-7.500***	-7.506***	-11.304***	-11.322***	-11.337***
DRI	Level	-2.400	-2.277	-0.644	-2.161	-2.066	-0.681
	First difference	-5.544***	-5.504***	-5.516***	-9.183***	-9.166***	-9.183***
GAS	Level	-3.034	-1.822	-0.398	-2.557	-1.571	-0.287
	First difference	-5.991***	-5.991***	-5.986***	-12.150***	-12.156***	-12.159***
MET	Level	-3.537	-3.252	-1.117	-3.697	-3.454	-1.316
	First difference	-8.872***	-8.891***	-8.907***	-7.545***	-7.568***	-7.591***
UREA	Level	-2.436	-2.155	-0.665	-3.059	-2.794	-0.874
	First difference	-9.114***	-9.132***	-9.148***	-8.593***	-8.619***	-8.648***

Note: The optimal number of lags for ADF is decided by the Schwarz Bayesian information criterion. The truncation lags for PP are decided by Newey–West default. Critical values are based on MacKinnon (1996). *** Significance at 0.01 level.

We conduct both unit tests on using the three different specifications (with intercept and trend, or with intercept, or without intercept and trend). Results are presented in Table C-3. The tests were first carried out in the levels of the variables and fail to reject the null hypothesis of non-stationarity. The tests were then carried out again in first differences and the results indicate that all of the individual series in first differences are stationary at the 1% significance level.

Based on these results, we proceed under the premise that all series can be fairly represented as I(1) stochastic processes. As a consequence, all regressions using the price level data, instead of first differences, will produce spurious estimation results. So, modeling will have to be done using first differences and the descriptive statistics for these first difference series are displayed in Table C-4.

Table C-4. Summary statistics for the first difference series

	ΔALU	ΔDIES	ΔDRI	ΔGAS	ΔMET	ΔUREA
Observations	241	241	241	241	241	241
Mean	-2.266	1.052	-0.105	0.019	0.644	0.58
Median	-3.166	2.276	-0.47	0	-0.396	1.501
Maximum	441.057	127.652	52.462	1.649	172.168	152.32
Minimum	-538.273	-204.802	-36.196	-2.981	-222.02	-293.771
Standard deviation	117.360	38.071	6.889	0.393	40.973	30.173
Skewness	-0.321	-0.926**	1.836***	-2.492***	-0.335	-3.521***
Excess Kurtosis	2.942*	6.246***	21.929***	23.053***	9.943***	42.524***
AD normality	2.0456***	5.4990***	21.7627***	25.7583***	19.1049***	18.9455***
LB(8)	28.716***	65.361***	102.06***	105.87***	127.34***	92.666***
LB(16)	34.722***	78.012***	142.99***	151.58***	144.57***	104.46***
$LB^2(8)$	64.287***	155.58***	26.852***	74.201***	69.781***	45.218***
$LB^2(16)$	68.008***	172.66***	58.346***	90.156***	77.661***	45.396***

Note: Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. AD is the Anderson-Darling test for the null hypothesis of normality, which is an improved version of the Kolmogorov–Smirnov test, where *** indicates non-normality at the 1% level (Anderson and Darling, 1952). LB (and LB^2) is the Ljung–Box Q-statistics computed on the first differences (squared first differences, respectively) where ** and *** indicate rejection of the no autocorrelation hypothesis at the 5% and 1% level respectively.

The distributional properties of these first difference series show some signs of non-normality. With the exception of DRI, the measures for skewness indicate that the monthly price difference series are negatively skewed. Beside, significant leptokurtosis is observed in all residual series apart from those associated with the aluminum market. Unsurprisingly, the goodness of fit test of Anderson and Darling (1952) systematically rejected the null hypothesis of distributional normality at the 1% level for all series. The Ljung–Box Q-statistics show signs of serial correlation in all series at different lag lengths. Furthermore, the Ljung–Box statistics computed on the squared residuals reveal clear evidence of volatility clustering or time dependent heteroscedasticity at lag length of up to 16. The joint presence of excess kurtosis and time-dependant heteroscedasticity is in favor of a model that incorporates some GARCH features.

Table C-5. Unconditional correlations among monthly first difference in prices

	ΔALU	ΔDIES	ΔDRI	ΔGAS	ΔMET	ΔUREA
ΔALU	1.000					
ΔDIES	0.419***	1.000				
ΔDRI	0.097	0.142**	1.000			
ΔGAS	0.020	-0.018	0.235***	1.000		
ΔMET	0.165**	0.067	-0.010	-0.013	1.000	
ΔUREA	0.203***	0.431***	0.124*	-0.018	0.170***	1.000
$\chi^2(15) = 128.421$ *(0.000)*						

Note: Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. $\chi^2(15)$ is the Bartlett statistic associated with a likelihood ratio test of the null hypothesis that the correlation matrix is equal to the identity matrix described in Morrison (1967, p. 113). The associated *p*-value is in brackets.

Table C-5 shows the unconditional correlations for the six first difference series. Again, the likelihood ratio test indicates that these correlations are, as a group, statistically significant. Interestingly, all the highly significant correlation coefficients displayed in this table are positive which may indicate the presence of co-movements. This finding is not surprising since these commodities are clearly related ones, and common macroeconomic shocks can affect the supply and demand of these products.

c - Cointegration

The I(1) nature of the price series calls for an investigation of the cointegration issue. Two or more time series may be non-stationary in levels (with a unit root), but a linear combination of these individual series may be stationary. If such a stationary linear combination exists, then the non-stationary time series are said to be cointegrated. The stationary linear combination is called a cointegrating equation and may be interpreted as a long-run equilibrium relationship between the variables. Any cointegration relationship will have to be taken into account when we formulate the model later on. If the six series are cointegrated, we may proceed with an error correction model in the equation and model the long- and short-run relationships among the variables using a Vector Error Correction Model (VECM). If they are not cointegrated, we simply specify these mean equations by using a standard vector-autoregressive (VAR) model.

We use the Johansen maximum likelihood approach employing both the maximum eigenvalue and trace statistic (Johansen, 1988; Johansen and Juselius, 1990). Prior to performing the cointegration

test, we have to determine the optimal number of lags. As the resulting model will be used to generate consistent scenarios of future prices, the optimal lag length has been chosen so as to minimize the Akaike's final prediction error (FPE) criterion.

Table C-6. Johansen and Juselius cointegration test

Null	Maximum eigenvalue		Trace statistic	
No. of CE(s)	Statistic	Critical Values (5%)	Statistic	Critical Values (5%)
None	54.64***	40.08	130.82***	95.75
At most 1	40.15***	33.88	76.18**	69.82
At most 2	17.21	27.58	36.02	47.86

Note: The selection of the lag length in the Johansen test procedure was based on Akaike's Final Prediction Error (FPE) criterion. r denotes the number of cointegrating vectors in the vector in the null hypothesis. Asterisks indicate rejection of the null hypothesis at the 5%** and 1%*** levels, respectively.

Table C-6 summarizes the results of the cointegration analysis between the six commodity prices. The test has been conducted using an unrestricted intercept specification. At the 5% level, both the maximum eigenvalue and the trace tests indicate that a co-integration rank of 2 is present. Therefore, we strongly reject the null hypothesis of no cointegration and assume the presence of two cointegrating relations.

Table C-7. Estimated cointegrating vectors

	ALU	DIES	DRI	GAS	MET	UREA
EC_1	1.000		1.174	76.444	-5.959	0.224
			[0.506]	*[2.270]*	*[-7.090]*	*[0.125]*
EC_2		1.000	-0.880	-55.917	-2.156	2.162
			[-1.031]	*[-4.509]*	*[-6.964]*	*[3.286]*

Note: t-statistics are in [ ]. These cointegrating vectors have been estimated using the Johansen procedure with a specification based on a lag length equal to 10, a value suggested by the FPE criterion.

The estimated cointegrating vectors are detailed in Table C-7. This finding reveals the presence of comovements in long term equilibrium adjustment which is not surprising given the "related" nature of the six commodities under scrutiny. At least three lines of explanations can be proposed for this phenomenon. Firstly, these commodities are far from being unrelated to one another as numerous supply-side linkages are at work. For example, both natural gas and crude oil (later refined in diesel oil) are jointly extracted from the so-called associated fields; natural gas is the main feedstock in the production of both urea and methanol; steel production is an energy intensive activities; both gas and electricity prices are tightly connected in some markets (Bunn, 2004) and this feature has some influence on aluminum smelting, which is a notorious electric intensive activity… Secondly, some specific institutional arrangements can also reinforce these linkages as in the case of the oil-product indexed pricing formulas implemented in most long-term gas import contracts (Asche et al., 2002). Lastly, common macroeconomic shocks (e.g., changes in industrial production, consumer prices, interest rates, stock prices or exchange rates) can affect the supply and demand of these commodities and thus explain the long-run behavior of their prices.

As we are dealing with cointegrated I(1) series, a Vector Error Correction Model (VECM) is required to model the long- and short-run relationships among the variables.

C.2 Methodology for an empirical model of commodity prices

Having examined the features of the time series above, we can now model the joint processes relating to monthly price changes for the six commodities. From the findings above, the use of a VECM is indicated to model the conditional mean equation. Besides, the descriptive statistics indicate the presence of ARCH effects and thus call for the implementation of a multivariate GARCH framework.

a - The conditional mean model

Let P_t denote a vector of m (here m=6) nonstationary prices $P_{i,t}$ at time t. Given the order of integration of the variables used, the data generating process of P_t can be appropriately modeled as a vector error correction model (VECM) with k−1 lags (which is derived from levels vector autoregression model (VAR) with k lags):

$$\Delta P_t = \mu + \Pi P_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta P_{t-j} + \varepsilon_t \qquad (C.1)$$

where Δ is the difference operator ($\Delta P_t = P_t - P_{t-1}$), μ is a $(m \times 1)$ constant vector, Γ_j is a $(m \times m)$ matrix of coefficients relating price changes lagged j periods to current changes in prices and thus describe the short run dynamics of the system. $\Pi = \alpha\beta'$ is a $(m \times m)$ matrix of coefficients relating lagged levels of prices to current changes in prices.[20] *The matrix Π can be decomposed into β which is the $(m \times r)$ cointegrating vector that determines the r long-term relationship(s) between the m series, and α which is the loading matrix that determines how the endogenous variables respond to disequilibrium in the long-run relationship(s). In case of r =0, the matrix Π will not be included in the model (it would thus be reduced to a standard VAR specification). Finally, ε_t is a $(m \times 1)$ vector of filtered residuals with a conditional variance-covariance matrix H_t such that $E(\varepsilon_t) = 0$ and $\operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$. This vector of unmodeled innovations reflects new information emanating from each of the series.*

b - A dynamic model of variances and covariances

A general multivariate GARCH framework may be defined as:

$$\varepsilon_t = H_t^{1/2} \eta_t \qquad (C.2)$$

[20] *Actually, Π may be of order m×(m+1) depending on whether the constant is inside or outside of the cointegration space.*

where η_t *is an i.i.d. vector error process of size* $(m \times 1)$ *named standardized residuals, such that* $E(\eta_t \eta_t^{'}) = I$.

Numerous multivariate GARCH models have been proposed. The list includes popular specifications such as the general VEC model (Bollerslev et al., 1988) and the BEKK-representation (Engle and Kroner, 1995) used in numerous empirical studies. These specifications are general but present a major burden for the present applied analysis: they impose the estimation of a large number of parameters. Given the limited size of our data set, the curse of dimensionality imposes us to look for a more parsimonious specification. Hopefully, the family of correlation multivariate GARCH models provides a convenient framework to model the dynamic processes of the variance-covariance matrix with a reduced number of parameters to be estimated. These models exploit the decomposition of conditional covariances into conditional standard deviations and conditional correlations. Under certain conditions, this decomposition reduces the computational complexity with separate estimation of the volatility and correlation parameters. Furthermore, these models guarantee the positive definiteness of the covariance matrix. Unsurprisingly, this convenient framework has been preferably used in numerous empirical studies.

Specifically, it assumes that the conditional correlation matrix can be expressed as:

$$H_t = D_t R_t D_t \qquad (C.3)$$

where $D_t = \mathrm{diag}\left\{\sqrt{h_{i,t}}\right\}$ *is a* $(m \times m)$ *diagonal matrix of time-varying standard deviations from univariate GARCH models and* $R_t = \left\{\rho_{ij,t}\right\}$ *is a possibly time-varying correlation matrix containing conditional correlation coefficients. The elements in* D_t *follow the univariate GARCH(* P *,* Q *) processes in the following manner:*

$$h_{i,t} = \omega_i + \sum_{p=1}^{P_i} \alpha_{i,p} \varepsilon_{i,t-p}^2 + \sum_{q=1}^{Q_i} \beta_{i,q} h_{i,t-q} \qquad \forall i = 1, \ldots, m. \qquad (C.4)$$

where $h_{i,t}$ *is the conditional variance of volatility of* $\varepsilon_{i,t}$ *for commodity* i *at time* t*,* ω_i *is a constant,* $\alpha_{i,p}$ *and* $\beta_{i,q}$ *are coefficients that are associated with the degree of innovation from lagged periods,* $\varepsilon_{i,t-p}^2$ *(ARCH term) and previous period's volatility spillover effects,* $h_{i,t-q}^2$ *(GARCH term) for each market respectively.*

In this paper, attention is focused on a particular specification within the family of correlation multivariate GARCH models: the Constant Conditional Correlation (CCC) model proposed by Bollerslev (1990). In the CCC model, a time invariant correlation matrix is assumed, i.e., $R_t = \overline{R}$*,* $\forall t$*. Thanks to this assumption, a consistent estimation can be obtained in two steps (Bollerslev,*

1990). The first step specifies univariate GARCH processes for the conditional variance of each series, $h_{i,t}$*. The constant conditional correlation is estimated using the standardized residuals* η_t *and the usual correlation estimator* $\overline{\rho}_{ij}$*. In the CCC model, the temporal variation in conditional covariance is thus solely determined by the time-varying conditional variance process. If the conditional variance process takes all positive values, and the correlation matrix* $\overline{R}$ *is positive definite, then the conditional covariance matrices is guaranteed to be positive definite.*

Of course, the constant correlation assumption plays a crucial role and thus deserves to be meticulously checked using, for example, the test presented in Engle and Sheppard (2001). Any empirical rejection of this hypothesis would obviously impose the use of an alternative specification that includes a time-varying conditional correlation matrix R_t*, such as, for example, the Dynamic Conditional Correlations (DCC) model proposed by Engle and Sheppard (2001) and Engle (2002).*

C.3 Estimation

Given the number of parameters to be simultaneously estimated and the relatively modest size of our data set (compared with those of a high frequency financial time series), difficulties in attaining convergence may be encountered when estimating the conditional mean, variance, and correlation equations in one single step. In contrast, the usual sequential estimation strategy is adopted. First, we estimate an appropriate VECM specification for the conditional mean equation and discuss its validity. Second, we focus the conditional variance-covariance matrix of the filtered residuals and estimate an appropriate multivariate GARCH model.

a - The conditional mean equation

According to the final prediction error (FPE) criterion, the optimal lag length k is equal to 10. Using this value and the obtained cointegrating rank (r = 2), the conditional mean equation can be estimated. To begin with, the cointegrating relationships are estimated using the Johansen procedure. As discussed above (cf. subsection C.1.c.), numerous sensible explanations can justify the existence of these long-run relationships among the prices of these related commodities. Using these two cointegrating vectors, the proposed VECM model has a total of 342 parameters, including the error correction and the intercept terms. Unsurprisingly, a number of estimated coefficients in this unrestricted model are statistically insignificant. Thus, we proceed to a model with a more parsimonious dynamic structure. We apply the Sequential Elimination of Regressors (SER) procedure presented in Brüggemann and Lütkepohl (2001). In this iterative "general-to-specific" procedure, the regressors with the lowest absolute t-statistics are successively eliminated provided that their absolute t-statistics do not exceed a threshold. This threshold is chosen such that the elimination procedure mimics model reduction on the basis of the Hannan-Quinn information criterion. Compared to the initial model, a total of 217 zero restrictions have been introduced in the parsimonious specification. The resulting model has been estimated using a feasible GLS procedure.

Estimation results are displayed in Table C-8 and Table C-9 and the diagnostics are presented in Table C-10. Interestingly a likelihood ratio test confirms that the restricted model is not rejected by the data: the joint test of all these restrictions gives a χ^2 statistic of 197.17 with a p-value equal to 0.829. The autoregressive structure of the estimated model seems to be statistically adequate, since the null hypothesis of no residual autocorrelation is never rejected by both the multivariate Breusch-Godfrey LM-test and the multivariate Portmanteau test. Hence, there is no evidence of un-modeled serial correlations. A multivariate Jarque-Bera test shows evidence of non-normality in the residuals. However, the performance of the maximum likelihood estimator of the cointegrating vectors is little affected by non-normal errors (Gonzalo, 1994). Regarding heteroscedasticity, a multivariate test for ARCH effect with three lags indicates the presence of significant ARCH effect in the residuals, and thus justifies the need to employ a GARCH framework.

Table C-8. Estimated coefficients for the parsimonious VECM (part A)

Equation									Adjusted R^2
ΔALU_t =	- 0.0789***	EC_1	+ 0.1684***	ΔALU_{t-1}	+ 3.2696***	ΔDRI_{t-1}	- 3.0596***	ΔDRI_{t-2}	0.301
	[-5.550]		*[3.053]*		*[3.116]*		*[-2.799]*		
	- 0.5443***	ΔMET_{t-2}	+ 0.5413**	ΔMET_{t-3}	- 0.8160***	$\Delta UREA_{t-3}$	+ 0.4578**	$\Delta DIES_{t-4}$	
	[-2.839]		*[2.579]*		*[-3.062]*		*[2.404]*		
	- 2.6109***	ΔDRI_{t-4}	- 46.1161**	ΔGAS_{t-4}	- 0.6226***	ΔMET_{t-4}	- 0.6301***	$\Delta UREA_{t-4}$	
	[-2.612]		*[-2.249]*		*[-3.178]*		*[-2.607]*		
	+ 0.7188***	$\Delta DIES_{t-5}$	+ 51.6089**	ΔGAS_{t-5}	+ 50.6490***	ΔGAS_{t-6}	- 0.8487***	$\Delta UREA_{t-6}$	
	[3.644]		*[2.413]*		*[2.604]*		*[-3.183]*		
	+ 33.8213*	ΔGAS_{t-7}	- 58.5713***	ΔGAS_{t-8}	+ 0.0985*	ΔALU_{t-9}	-1.1533***	$\Delta UREA_{t-9}$	
	[1.932]		*[-2.811]*		*[1.811]*		*[-4.030]*		
	+ 88.2971***								
	[5.050]								
$\Delta DIES_t$ =	0.1873***	$\Delta DIES_{t-1}$	+ 0.8790***	ΔDRI_{t-1}	+ 11.5631*	ΔGAS_{t-1}	+ 0.1410**	$\Delta DIES_{t-2}$	0.325
	[3.289]		*[2.788]*		*[1.803]*		*[2.391]*		
	- 0.3510***	$\Delta UREA_{t-3}$	- 18.2514***	ΔGAS_{t-4}	+ 13.0386**	ΔGAS_{t-5}	- 0.1932***	$\Delta UREA_{t-5}$	
	[-4.914]		*[-2.767]*		*[2.364]*		*[-2.621]*		
	- 0.0386**	ΔALU_{t-6}	- 11.3934**	ΔGAS_{t-6}	+ 0.1299**	ΔMET_{t-6}	+ 0.1321**	ΔMET_{t-7}	
	[-2.207]		*[-2.061]*		*[2.280]*		*[2.325]*		
	- 0.3487***	$\Delta UREA_{t-7}$							
	[-4.485]								
ΔDRI_t =	- 0.0023***	EC_1	+ 0.3820***	ΔDRI_{t-1}	- 0.0385***	ΔMET_{t-1}	+ 0.0596***	$\Delta UREA_{t-1}$	0.529
	[-3.492]		*[7.159]*		*[-4.551]*		*[4.772]*		
	+ 0.0288***	$\Delta DIES_{t-2}$	+ 3.9880***	ΔGAS_{t-2}	- 0.0433***	$\Delta UREA_{t-2}$	+ 0.0320***	$\Delta DIES_{t-3}$	
	[2.881]		*[4.042]*		*[-3.096]*		*[3.430]*		
	+ 0.1126**	ΔDRI_{t-4}	- 0.0269***	ΔMET_{t-5}	- 0.1944***	ΔDRI_{t-6}	+ 0.0384***	$\Delta UREA_{t-6}$	
	[2.176]		*[-3.018]*		*[-3.712]*		*[2.921]*		
	-0.0297**	$\Delta UREA_{t-7}$	+ 0.1523***	ΔDRI_{t-8}	- 0.1245**	ΔDRI_{t-9}	+ 2.6147***		
	[-2.362]		*[2.750]*		*[-2.336]*		*[3.1506]*		

Note: t-statistics are in [ ]. Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. EC_1 and EC_2 are the two cointegrated combinations detailed in Table C-7.

Table C-9. Estimated coefficients for the parsimonious VECM (part B)

Equation									Adjusted R^2
ΔGAS_t =	0.0015***	$\Delta DIES_{t-1}$	- 0.4545***	ΔGAS_{t-1}	- 0.0013**	$\Delta UREA_{t-1}$	+ 0.0024***	$\Delta DIES_{t-2}$	0.792
	[4.178]		*[-8.857]*		*[-2.358]*		*[6.370]*		
	- 0.2406***	ΔGAS_{t-2}	+ 0.0009***	ΔMET_{t-2}	- 0.0024***	$\Delta UREA_{t-2}$	- 0.0002*	ΔALU_{t-3}	
	[-4.730]		*[3.059]*		*[-4.474]*		*[-1.931]*		
	+ 0.0015***	$\Delta DIES_{t-3}$	+ 0.2712***	ΔGAS_{t-3}	+ 0.0023***	$\Delta UREA_{t-3}$	+ 0.0040***	$\Delta DIES_{t-4}$	
	[3.412]		*[5.821]*		*[4.042]*		*[10.221]*		
	+ 0.0110***	ΔDRI_{t-4}	+ 0.0012**	$\Delta UREA_{t-4}$	+ 0.0003***	ΔALU_{t-5}	+ 0.0025***	$\Delta DIES_{t-5}$	
	[5.366]		*[2.026]*		*[3.022]*		*[5.333]*		
	+ 0.0028***	$\Delta DIES_{t-6}$	+ 0.0022***	$\Delta DIES_{t-7}$	- 0.0024***	$\Delta UREA_{t-7}$	+ 0.0011**	$\Delta DIES_{t-8}$	
	[6.213]		*[4.679]*		*[-4.113]*		*[2.455]*		
	+ 0.0023***	$\Delta UREA_{t-8}$	+ 0.0002*	ΔALU_{t-9}	+ 0.0012***	$\Delta DIES_{t-9}$	- 0.1094***	ΔGAS_{t-9}	
	[4.259]		*[1.947]*		*[2.683]*		*[-3.410]*		
	- 0.0029***	$\Delta UREA_{t-9}$							
	[-5.534]								
ΔMET_t =	0.0054***	EC_1	+ 0.0284***	EC_2	+ 16.8387**	ΔGAS_{t-1}	+ 0.7749***	ΔMET_{t-1}	0.583
	[2.906]		*[3.338]*		*[2.444]*		*[14.105]*		
	- 0.0334**	ΔALU_{t-2}	- 0.3448***	ΔMET_{t-2}	+ 0.4065***	ΔMET_{t-3}	- 0.1577**	$\Delta UREA_{t-3}$	
	[-2.135]		*[-5.120]*		*[5.716]*		*[-2.010]*		
	+ 0.0736***	ΔALU_{t-4}	- 0.1110*	$\Delta DIES_{t-4}$	- 0.6162**	ΔDRI_{t-4}	- 20.4702***	ΔGAS_{t-4}	
	[4.415]		*[-1.853]*		*[-2.098]*		*[-3.563]*		
	- 0.3592***	ΔMET_{t-4}	- 0.3170***	$\Delta UREA_{t-4}$	0.2842***	$\Delta DIES_{t-5}$	- 17.7523***	ΔGAS_{t-5}	
	[-5.357]		*[-3.985]*		*[4.504]*		*[-2.810]*		
	+ 0.1462**	ΔMET_{t-5}	- 0.1656**	$\Delta UREA_{t-6}$	+ 0.1887***	ΔMET_{t-7}	+ 17.6121***	ΔGAS_{t-8}	
	[2.453]		*[-2.118]*		*[3.226]*		*[3.176]*		
	- 0.1811***	ΔMET_{t-8}	+ 0.0544***	ΔALU_{t-9}	+ 0.2005**	$\Delta UREA_{t-9}$			
	[-3.316]		*[3.128]*		*[2.508]*				
$\Delta UREA_t$ =	0.0201***	EC_1	- 0.0670***	EC_2	+ 0.1667***	$\Delta DIES_{t-1}$	+ 0.8383***	ΔDRI_{t-1}	0.653
	[5.236]		*[-7.054]*		*[4.489]*		*[4.265]*		
	+ 0.1008***	ΔMET_{t-1}	+ 0.2364***	$\Delta UREA_{t-1}$	+ 0.2121***	$\Delta DIES_{t-2}$	- 15.6010***	ΔGAS_{t-2}	
	[3.062]		*[5.166]*		*[5.764]*		*[-3.638]*		
	+ 0.5893***	ΔDRI_{t-3}	- 0.2749***	$\Delta UREA_{t-3}$	+ 0.0701*	$\Delta DIES_{t-4}$	- 0.8140***	ΔDRI_{t-4}	
	[2.999]		*[-6.086]*		*[1.860]*		*[-3.888]*		
	- 0.0266**	ΔALU_{t-6}	+ 0.1274***	$\Delta DIES_{t-6}$	+ 1.3829***	ΔDRI_{t-6}	- 0.0610*	ΔMET_{t-6}	
	[-2.368]		*[3.176]*		*[6.954]*		*[-1.756]*		
	+ 0.0722*	$\Delta DIES_{t-7}$	- 0.1858***	$\Delta UREA_{t-7}$	- 0.0567***	ΔALU_{t-8}	+ 0.1091***	$\Delta DIES_{t-8}$	
	[1.944]		*[-3.762]*		*[-5.200]*		*[3.001]*		
	- 0.3644**	ΔDRI_{t-8}	+ 0.1402***	ΔMET_{t-8}	- 0.0444***	ΔALU_{t-9}	+ 0.0686*	$\Delta DIES_{t-9}$	
	[-2.033]		*[3.938]*		*[-3.902]*		*[1.798]*		
	- 9.7102***	ΔGAS_{t-9}	+ 0.1044***	ΔMET_{t-9}	- 38.4993***				
	[-3.112]		*[2.994]*		*[-6.0269]*				

Note: t-statistics are in [ ]. Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. EC_1 and EC_2 are the two cointegrated combinations detailed in Table C-7.

The diagnostic statistics also show that the model explains relatively large proportions of the variations in prices. The adjusted R^2 values ran from 0.301 (aluminum) to 0.792 (natural gas). These values suggest that the models fit the data quite well.

Table C-10. Diagnostic checks of the conditional mean equations

Multivariate tests		Statistic	*p*-value
Normality:	JB, $\chi^2(12)$	2222.58	*(0.000)*
Autocorrelation:	LM(10), $\chi^2(360)$	377.26	*(0.255)*
	LM(12), $\chi^2(432)$	462.70	*(0.148)*
	LB(14), $\chi^2(382)$	378.15	*(0.546)*
	LB(24), $\chi^2(742)$	742.46	*(0.488)*
Presence of ARCH:	ARCH(3), $\chi^2(1323)$	1911.86	*(0.000)*

Note: JB is Jarque-Bera multivariate statistic based on Doornik and Hansen (2008). All the other tests are those described in Lütkepohl (2005): LM(x) is the multivariate Breusch-Godfrey LM-test for the xth order autocorrelation, LB(y) is the multivariate portmanteau test for residual autocorrelation up to the order y, and ARCH(3) is the multivariate LM-test for ARCH effect with 3 lags.

As our main objective is the construction of a data-driven price forecasting model, the economical interpretation of these empirical findings will not be developed extensively. However, two points deserve a discussion. Firstly, attention is drawn to the restrictions imposed on the adjustment coefficients (i.e. on the loading matrix α). According to the results obtained with the SER procedure, zero restrictions are simultaneously imposed on all the loading coefficients in both diesel oil and natural gas equations. It means that the prices of these two commodities are able to influence the long-run stochastic paths of the other four commodities without being themselves influenced by the long-run paths of these commodities. Given the importance of this assumption for the adjustment structure, the validity of this restriction must be meticulously checked. That's why we apply the Likelihood Ratio (LR) test proposed by Johansen on the unrestricted VECM (Johansen and Juselius, 1990; Johansen, 1995). Interestingly, the null hypothesis of long-run weak exogeneity is rejected for all commodities but diesel oil and natural gas (see Table C-11). Furthermore, a joint test shows that the combined long-run weak exogeneity of these two fuels is not rejected by the data. These zero restrictions are thus justified.

Table C-11. Long-run weak exogeneity tests

	ALU	DIES	DRI	GAS	MET	UREA	joint test for { DIES , GAS }
Statistic	14.328***	**3.380**	6.866**	**0.092**	8.535**	36.631***	**3.495**
p-value	*(0.000)*	***(0.185)***	*(0.032)*	***(0.955)***	*(0.014)*	*(0.000)*	***(0.479)***

Note: The table displays the Likelihood Ratio (LR) statistics computed on the unrestricted VECM (Johansen and Juselius, 1990; Johansen, 1995). Under the null hypothesis that the series does not respond to perturbations in the long-run relationships (i.e. zero restrictions on the loading coefficients), the statistics are distributed as chi-squared with degrees of freedom equal to the number of zero restrictions in the loading vector (i.e. 2 in all cases but the joint test that requires 4 degrees of freedom). Asterisks indicate rejection of the null hypothesis at the 5%** and 1%*** levels, respectively and the use of bold indicates no rejection at the 10% level.

Secondly, we focus on the short-run dynamics estimates and notice that all six markets exhibit significant own mean spillovers. Interestingly, these own mean spillovers intervene with a delay in all cases but aluminum. Besides, significant cross spillovers are also observed indicating that lagged price differences in the other markets can be used to forecast the price movements of a given

commodity. Such a finding is not surprising given the large supply and demand relationships that exist between these commodities.

b - The conditional variance equation

The estimation of the proposed specification is based on the maximum likelihood method and involves the algorithm presented in Berndt et al. (1974). As some signs of non-normality are present in the residual series, we employ Bollerslev and Wooldridge's (1992) quasi-maximum likelihood method to generate consistent standard errors that are robust to non-normality.

To keep the model parsimonious, the conditional variances are modeled using a low order GARCH specification: $P_i = Q_i = 1$, $\forall i = 1,...,m$. *As the assumption of a time-invariant conditional correlation matrix plays a crucial role, the test for constant correlation* $R_t = \bar{R}$ *due to Engle and Sheppard (2001) has been conducted. Several specifications have been used for the alternative hypothesis*[21] *but the tests systematically failed to reject the null hypothesis of a constant correlation matrix. Hence, we proceed using the CCC specification.*

From the estimation results reported in Table C-12, several facts stand out. First, very high level of significance are attached to most of the estimated coefficients for the lagged variance β_i. *These coefficients represent the own lagged volatility spillovers. There are also significant coefficients for the own-innovation spillover effect (*α_i*). These remarks justify the appropriateness of the GARCH(1,1) specification. The relative magnitudes of the estimated coefficients show that the own-innovation spillover effect (*α_i*) is dominated by the lagged volatility spillover effect (*β_i*) in all cases. In all equations,* α_i *and* β_i *sum up to a number less than one, which is required in order to have a mean reverting variance process. However, for some commodities such as diesel oil, DRI and methanol, these sums are larger than 0.90, implying that the volatility displays a rather high persistence.*

As the null hypothesis of a correlation matrix equal to the identity matrix is firmly rejected, some second moment linkages are at work between these commodity markets which confirms the pertinence of a multivariate GARCH approach. Looking at the correlation matrix, we easily identify the presence of significant values for some of the off-diagonal coefficients, indicating that lagged innovations in one commodity market spill over into the variance observed in other markets: as, for example between aluminum and diesel oil, or between DRI and natural gas.

---

[21] *The alternative hypothesis is:* $vech(R_t) = vech(\bar{R}) + \rho_1 vech(R_{t-1}) + ... + \rho_n vech(R_{t-n})$. *The tests have been conducted with lag lengths varying from 1 to 12. To save space the results of these tests are not reported here but are available from the authors upon request.*

Table C-12. CCC-GARCH model estimates and diagnostic test results

	ΔALU $i = 1$	ΔDIES $i = 2$	ΔDRI $i = 3$	ΔGAS $i = 4$	ΔMET $i = 5$	ΔUREA $i = 6$
Panel A – GARCH estimates						
ω_i	1 624.3913***	13.3677	3.2444	0.0041*	41.5661**	30.9672**
	[2,7007]	*[0,7684]*	*[1,0310]*	*[1,6522]*	*[2,1023]*	*[2,0530]*
α_i	0.1690**	0.0604*	0.3756	0.1140**	0.2759**	0.2134***
	[2,2745]	*[1,8992]*	*[1,3137]*	*[2,0476]*	*[2,0321]*	*[2,7428]*
β_i	0.6133***	0.9256***	0.5560*	0.7213***	0.6826***	0.6835***
	[5,8337]	*[18,7470]*	*[1,8542]*	*[5,9487]*	*[7,5427]*	*[6,8509]*
Panel B – Correlation matrix estimates and related diagnostics						
$\overline{\rho}_{2i}$	0.2388***					
	[4.2246]					
$\overline{\rho}_{3i}$	0.0563	0.0027				
	[0.8492]	*[0.0431]*				
$\overline{\rho}_{4i}$	0.0635	0.1084	-0.1550**			
	[0.9905]	*[1.4575]*	*[-2.3848]*			
$\overline{\rho}_{5i}$	0.1461**	0.0239	-0.0272	-0.0483		
	[2.4365]	*[0.3443]*	*[-0.4792]*	*[-0.7288]*		
$\overline{\rho}_{6i}$	0.015	0.2627***	-0.0667	0.0886	0.0429	
	[0.2432]	*[4.4219]*	*[-0.9751]*	*[1.2901]*	*[0.5592]*	
Associated test:	χ^2 (15) = 48.3461 *(0.000)*					

Note: z-statistics based on robust standard errors are in [ ]. Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. $\chi^2(15)$ is the Bartlett statistic associated with a likelihood ratio test of the null hypothesis that the correlation matrix is equal to the identity matrix described in Morrison (1967, p. 113). The associated *p*-value is in brackets.

Before exploiting this empirical model, one should test to see if its appropriateness by using a series of diagnostics tests (cf. Table C-13). Our low order GARCH(1,1) specification satisfactorily models the second moments dynamics since both the LM test for ARCH, and the Ljung-Box Q-statistics on the squared standardized residuals show no signs of un-modeled GARCH effects in the residuals. Besides which, we use the BDS procedure (Brock et al., 1996) to test the null hypothesis that the time series under consideration is generated by identically and independently distributed (i.i.d.) stochastic variables. Possible causes of rejection of the i.i.d. assumption are un-modeled non-linear dependences or the existence of a chaotic structure embedded in series (Hsieh, 1991). If evidence of nonlinearity is still found in the standardized residuals, this must cast doubt on the model's adequacy. Considering the test statistics reported in Table C-13, the i.i.d. assumption can not be rejected at the 10% significance level. This finding suggests that the estimated GARCH model effectively explains the non-linearities present in the VECM residuals as there is no remaining forecastable structure embodied within the standardized residual series. In addition, we can investigate the distributional properties of the standardized residual series. These series show acceptable signs of multivariate normality as the Anderson and Darling tests failed to reject normality at the 10% level in all series but one (DRI). Thus, we proceed assuming that the standardized residuals are i.i.d. and are normally distributed.

Table C-13. Diagnostics tests conducted on the standardized residuals

	ΔALU $i = 1$	ΔDIES $i = 2$	ΔDRI $i = 3$	ΔGAS $i = 4$	ΔMET $i = 5$	ΔUREA $i = 6$
$LB^2(12)$	10.538	5.3951	2.9349	3.7013	5.1015	8.9767
ARCH-LM(4)	2.5267	1.0091	0.6790	0.5832	2.3657	2.1364
BDS(2)	-1.3810	0.4652	-1.0617	1.5250	-1.2621	-0.4803
AD normality	0.2465	0.7803	3.1324**	1.1023	1.1170	0.3223

Note: Asterisks indicate rejection of the null hypothesis at the *10%, ** 5% and *** 1% level respectively. $LB^2(x)$ is the Ljung–Box Q-statistics computed on the squared standardized residuals, where asterisks indicate rejection of the no autocorrelation up to the order x hypothesis. ARCH-LM(y) is the usual Engle's LM test with y lags for the null hypothesis that a series exhibits no ARCH effects (Engle, 1982). BDS(m) is the non-parametric test for the null hypothesis of the i.i.d. assumption of Brock et al. (1996) for the embedding dimension m and a bound ε equal to 1/2 standard deviation for each series. Following Brock et al. (1993) who suggest the use of bootstrapping when applying the BDS test to standardized residuals from GARCH models, rejection is evaluated on the basis of the bootstrapped p-values. AD is the Anderson-Darling test for the null hypothesis of normality (Anderson and Darling, 1952).

C.4 Concluding remarks on this empirical model

This empirical model highlights the presence of significant spillover effects in both mean and variance among the different commodities. Arguably, the commodities that exhibit both mean and variance cross spillovers are more strongly linked than the others. Several lines of explanations can be proposed for this phenomenon and the list includes: (i) the existence of supply-side linkages, (ii) the role played by institutional arrangements such as the oil-product indexed pricing formulas implemented in most long-term gas import contracts, (iii) the possible influence of common macroeconomic shocks and (iv) the possible "excess co-movement" i.e., price movements that are not accounted for by the common effects of exogenous macroeconomic variables.

According to the results of the diagnostic checks, this time series model is a well-behaved one. We thus proceed assuming that this parsimonious VECM-GARCH is satisfactorily a close approximation to the actual data generating process. In our study of export diversification, we used Monte Carlo simulations of this empirical model to generate a large number of possible future prices paths.