

Dipartimento di Impresa e Management

Cattedra di Statistica

**L'influenza delle esportazioni sul PIL
pro-capite italiano**

RELATORE
Prof.ssa De Giovanni

CANDIDATO
Vito Guerra
Matr. 172891

Anno accademico 2012-13

INDICE

Introduzione	1
1. Primo capitolo.....	3
1.1. La situazione economica italiana tra “luci ed ombre”.....	3
1.2. La crisi del “debito sovrano”	4
1.3. Rapporto Annuale ISTAT 2013: capitolo II	6
1.4. La crisi dalla prospettiva delle imprese italiane	7
1.5. Le strategie delle imprese italiane in risposta alla contrazione.....	8
1.6. Le esportazioni e il “Made in Italy”	9
1.7. Analisi dell’export settoriale nel 2011-2012.....	9
1.8. Analisi delle variabili del modello di regressione.....	12
2. Secondo Capitolo.....	14
2.1. Il modello di regressione	14
2.2. Le variabili: dipendenti ed indipendenti.....	14
2.3. Il modello di regressione come esemplificazione della realtà	15
2.4. Il metodo dei minimi quadrati.....	18
2.5. ... E le sue ipotesi.....	19
2.6. Modello di regressione multivariato	21
2.7. Il coefficiente di determinazione.....	25
3. Terzo Capitolo	27
3.1. Analisi di regressione: due casi estremi	30
3.2. Modello di regressione selezionato mediante metodo “best subset” ..	33
3.3. Il modello	34
3.4. Bontà di adattamento del modello ai dati.....	38
3.5. Oltre la bontà di adattamento del modello ai dati	41
4. Osservazioni e conclusioni	42
5. Ringraziamenti	48
6. Bibliografia	49

Introduzione

Ognuno di noi , negli ultimi anni, ha dovuto fronteggiare nel suo piccolo le conseguenze, per chi catastrofiche, per chi semplicemente negative (nell'accezione più ampia del suo significato) della crisi economica mondiale sviluppatasi a partire dal 2008.

A livello aggregato, le conseguenze di questo terremoto le sta vivendo tutt'ora l'Italia sotto forma di recessione: l'obiettivo di questo lavoro, una volta analizzate le determinanti del PIL italiano, in termini di valore aggiunto per il singolo cittadino (PIL Pro-capite) , è quello di illustrare con particolare enfasi quella che, dalla ricerca effettuata, risulta essere stata la principale fonte d'ossigeno di cui l'Italia ha beneficiato negli ultimi anni in termini di competitività all'estero: le esportazioni.

Mediante analisi di regressione lineare multipla, verrà infatti dimostrato come "lo scambio di merci italiane con il mondo" sia stata la variabile determinante a sostegno del pil-procapite italiano.

L'analisi spazierà, partendo dalla situazione economica italiana attuale, passaggio necessario per comprendere quelle che sono le questioni di natura strutturale che verranno affrontate, al vero e proprio calcolo statistico delle variabili che saranno prese in considerazione per la stima della significatività delle esportazioni (e di altre variabili) rispetto al pil procapite.

Così come l'analisi del passato è necessaria alla comprensione del presente, altrettanto importante è l'analisi del presente, per poter riuscire a prevedere quelle che saranno le dinamiche future.

Ed è proprio per questo motivo che dopo una premessa sui metodi statistici d'analisi che verranno richiamati all'interno del secondo capitolo, sarà necessaria l'analisi puntuale dei dati forniti dall' ISTAT;

dati, che saranno applicati all'interno di un modello di regressione lineare multivariata.

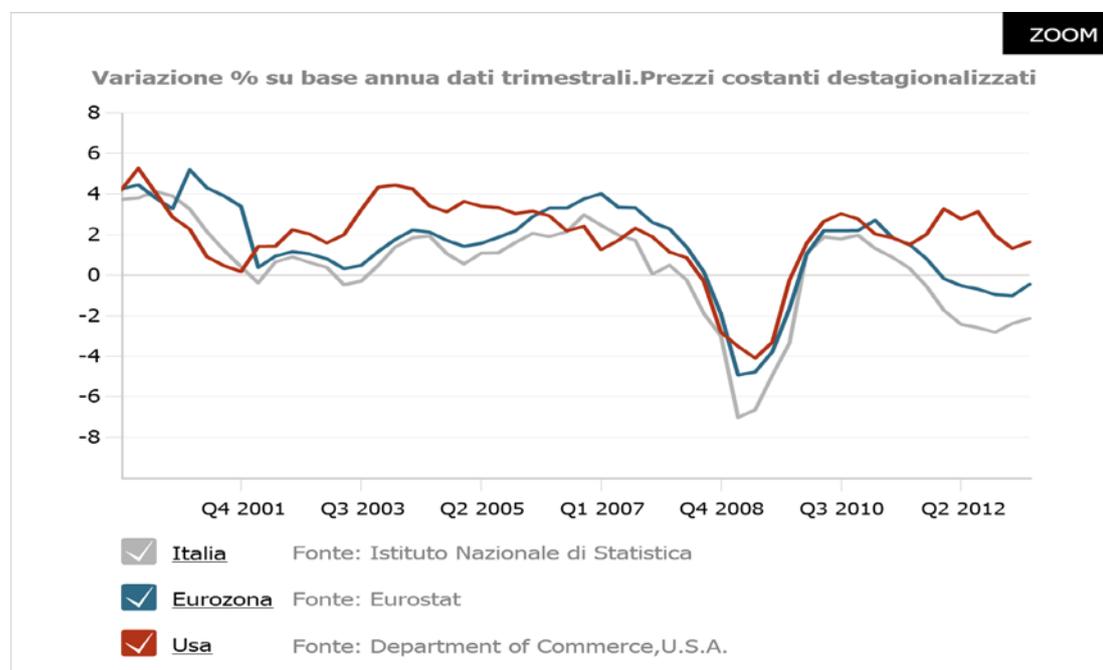
In ultima battuta, grazie all'ausilio delle stime ottenute mediante la regressione, sarà possibile trarre delle conclusioni circa l'andamento dell'economia italiana nell'orizzonte futuro di breve periodo.

In primis consideriamo brevemente la situazione economica italiana odierna.

Capitolo 1

LA SITUAZIONE ECONOMICA ITALIANA TRA LUCI ED OMBRE:

La violenta crisi economica che ha investito negli ultimi anni le economie di tutto il mondo, sta gradualmente lasciando il passo alla ripresa in buona parte dei Paesi che ha tramortito. Questo grazie all'impegno ed alla reazione dei governi e delle popolazioni messe a dura prova dal recente cataclisma. Purtroppo in Italia la ripresa stenta a decollare e l'economia permane ad oggi ancora in uno stato recessivo. Dai dati EUROSTAT emersi a settembre 2013, si evince come "l'Eurozona" sia nel complesso finalmente riuscita a mettere il segno positivo dinanzi al suo Prodotto Interno Lordo: lo stesso dovrebbe accadere nel 2014 secondo le proiezioni del medesimo istituto di statistica. Nel seguente grafico, vengono riportati gli andamenti dei PIL rispettivamente italiano, europeo e statunitense. (grafico fonte [IlSole24Ore](#))



Come si evince dall'immagine, i saggi di crescita europei fino al secondo quadrimestre del 2012 (ma il trend non è mutato successivamente) si

attestavano su valori superiori a quelli italiani, il che significa che l'Italia è cresciuta meno della media europea, laddove c'era crescita.

L'Europa nel complesso però, rappresenta un "Leviatano" al cui interno sono presenti numerose realtà molto differenti tra loro: sarebbe infatti estremamente riduttivo analizzare l'economia europea nel complesso, dimenticandosi delle singole realtà nazionali che compongono la stessa unione economica. Occorre considerare dunque, gli ultimi terremoti finanziari che hanno scosso l'UE, e nella fattispecie la crisi del "Debito Sovrano".

LA CRISI DEL "DEBITO SOVRANO":

La crisi del debito sovrano ha minato pesantemente la credibilità europea nel mondo, laddove sono emerse via via delle crepe all'interno del sistema economico di diversi stati: Irlanda, Portogallo, Spagna, Grecia, e da ultima, non per importanza, l'Italia. La crisi del debito, è stata senz'altro la causa principale per cui si è reiterata negli ultimi due anni la stagnazione e recessione dei singoli Paesi, e dell'Eurozona nel complesso.

Si sono resi necessari sforzi non indifferenti per far sì che i Paesi membri dell'unione riuscissero a rispettare i dettami di cui è stata inderogabilmente definita l'obbedienza, nella fattispecie in cui le finanze pubbliche risultavano dissestate.

La risposta alla situazione d'emergenza, ha trovato "nell'austerità" il suo minimo comune denominatore, e nelle politiche fiscali restrittive il suo cavallo di battaglia. E così gradualmente, grazie alle suddette manovre finanziarie, ed ad altre operazioni di rilevante importanza Macroeconomica circa la politica monetaria (poste in essere dal **Fondo europeo di stabilità finanziaria** e dal **Meccanismo europeo di stabilità**), gradualmente le singole economie europee sono riuscite a guadagnare sempre più fiducia: lo spread, ossia il differenziale tra titoli di stato, si è ridotto. I mercati quindi hanno percepito

relativamente meno rischiose le economie europee un tempo considerate “malate”.

Passando alla condizione attuale, possiamo considerare come essa presenti sia spiragli di luce, ma al contempo numerose ombre che continuano a perturbare il panorama economico.

Concentrando il focus interamente sulla situazione italiana, emerge un quadro ancora piuttosto enigmatico: sembrano essere passati anni luce dalle settimane terribili in cui lo spread schizzava ad oltre 570 punti base, ed in cui la credibilità italiana era quasi nulla a causa anche della crisi del IV Governo Berlusconi. In quel momento storico i mercati smettevano quasi definitivamente di scommettere sul riassetto dell'economia italiana.

Sebbene la situazione odierna conceda molto più ossigeno rispetto a quel lasso di tempo, paghiamo ancora lo scotto della crisi del debito sovrano sotto forma di recessione.

Dalla flessione del PIL del 2,4% nel 2012, la proiezione più attendibile per l'anno corrente sembra essere secondo l'istituto nazionale di statistica ISTAT una flessione dell'1,8%. Il trend, per quanto possa consolare, risulta positivo, il che sembrerebbe presagire che il peggio, ad ora, sia passato.

Possiamo quindi considerare le “ombre”, come gli spettri recessivi, che ancora per il momento aleggiano sul nostro Paese, ma che sembrerebbe si stiano lentamente diradando, ponendo come orizzonte temporale per la ripresa anche il prossimo anno.

Infatti, pur analizzando le serie storiche degli ultimi trimestri, che rilevano delle percentuali di segno negativo del 2,2% nel primo trimestre 2013 e del -1% nel secondo, la proiezione più accreditata per il 2014 sembrerebbe essere di segno positivo: **+0,7%! (dati fonte ISTAT)** Non va commesso però l'errore di far coincidere l'uscita dalla recessione con la cessazione di tutti i problemi: la

struttura della nostra economia, ancora fragile, non ci consente infatti un consolidamento stabile e duraturo.

La ripresa quindi, (se) spiccherà il volo, lo farà solo ed esclusivamente grazie alle esportazioni, unico ed intramontabile “motore dell’economia italiana”. Non grazie ai consumi, alla spesa pubblica o agli investimenti, chi più chi meno tutti ancora in flessione.

In un contesto economico così incerto e frammentato tra “luci ed ombre”, segnali di ripresa e campanelli d’allarme, una volta considerato lo scenario economico attuale, spostiamo il focus sul settore industriale italiano, analizzato recentemente nella relazione annuale 2013.

RAPPORTO ANNUALE ISTAT 2013, capitolo II:

L’importanza di questo documento risulta a tratti imprescindibile, laddove il filo del discorso tocca le variabili macroeconomiche di cui ci occuperemo; in esso, ed in particolare all’interno del capitolo secondo, rileveremo alcuni dati e parametri necessari ad affrontare il tema delle esportazioni italiane e del loro impatto sul valore aggiunto del nostro Paese.

Passeremo quindi brevemente in rassegna le tematiche più prettamente qualitative analizzate nel documento, per poi concentrarci sulle diverse tipologie di variabili che indicheremo come le “determinanti” del nostro modello.

Partendo dai problemi strutturali che interessano il settore della produzione industriale italiana, consideriamo una serie di elementi utili a comprendere l’enorme impatto che è stato sostenuto negli ultimi anni di crisi dalla classe imprenditoriale italiana in senso lato.

LA CRISI ECONOMICA DALLA PROSPETTIVA DELLE IMPRESE ITALIANE:

Le industrie italiane si sono trovate costrette, e si trovano tutt'ora, a fronteggiare numerosi ostacoli di diversa natura: istituzionale, creditizia, burocratica e più generalmente strutturale.

Il fattore che più di tutti ha inciso in negativo sulla competitività delle imprese italiane, differentemente da quanto molti potrebbero essere tentati a pensare, è stata la scarsità di fonti di finanziamento per alimentare gli investimenti e per riallineare l'economia reale durante la crisi.

Il dato è sconcertante: il 40% del gap di competitività di cui hanno pagato dazio le nostre imprese è derivante, appunto, da problemi di natura burocratica ed amministrativa, e dalla difficoltà di reperire, laddove ce ne fossero, risorse finanziarie.

Tale mancanza, la quale incide quasi per la metà del gap di competitività perso negli ultimi anni, supera persino il calo della domanda interna, uno dei fattori che tradizionalmente accompagna il circolo vizioso che si innesta laddove l'economia reale è depressa, connotandosi come principale elemento della congiuntura negativa.

E sembrerebbe tutto fuorchè consolante arrivare alla conclusione che non è stata la mancanza di infrastrutture e/o il picco negativo della domanda interna a far precipitare la competitività dell'economia italiana, ma è per colpa della chiusura "dei rubinetti", quindi della crisi del credito (la quale ha osteggiato l'accesso al credito) che oggi il nostro Paese abbia perso punti percentuali in doppia cifra a livello di produttività industriale.

LE STRATEGIE DELLE IMPRESE ITALIANE IN RISPOSTA ALLA CONTRAZIONE:

Il nostro sistema economico, com'è conosciuto nel mondo, è quello del capitalismo detto a "conduzione familiare" (70% dell'intero settore industriale). E, laddove la conduzione delle imprese segue l'impronta più propriamente familiare, le strategie poste in essere per fronteggiare le emergenze senz'altro si baseranno sulla difesa piuttosto che l'attacco: non c'è da meravigliarsi quindi, che nel 2011 (Dati ISTAT), due imprese su tre abbiano attuato tali misure strategiche (difensive).

Tra i meccanismi difensivi posti in essere riscontriamo in primis la difesa della propria quota di mercato. Come contraltare, specialmente tra le grandi imprese, la strategia offensiva ha riscontrato molto più successo in termini percentuali: oltre a quest'orientamento "espansivo" è stato rilevato un ultimo meccanismo di ristrutturazione strategica, ossia l'implementazione di una rete sempre più fitta di infrastrutture, distretti e joint ventures.

Queste possono essere considerate le strategie attuate dalle imprese italiane per cercare di superare lo scoglio della crisi. Ognuna di esse, però contiene al suo interno un principio, che va ben al di là dell'orizzonte delle semplici istanze strategiche e può essere considerato il vero movente della possibile rinascita italiana: la qualità del prodotto.

LE ESPORTAZIONI ed il “MADE IN ITALY”

Nella tradizione di innata eccellenza del cosiddetto “Made in Italy”, che ha caratterizzato fortunatamente sempre in positivo la nostra economia, ben due imprese su tre, per uscire dalla crisi hanno letteralmente “scommesso” sulla qualità del proprio prodotto. Il restante 33% è rimasto piantato alle politiche di costo che come abbiamo precedentemente ricordato, viene ad essere annoverato all’interno delle strategie difensive.

Ad essere più precisi, il range entro il quale si assestano le percentuali di imprese italiane che hanno puntato sulla qualità del prodotto e sull’ampliamento dell’offerta al mercato oscillano tra il 67% ed il 76%, a seconda che si considerino le fasce in cui di solito vengono divise le imprese, in base al criterio dimensionale (Piccole, Medie, Grandi imprese).

Ed è proprio grazie a questa ultima accezione, ossia alla continua ricerca della qualità doviziosa al minimo dettaglio, che i nostri prodotti sono riusciti a continuare ad imporsi all’estero, ed a far rifiatore sia il PIL che la bilancia commerciale italiana.

ANDAMENTO Dell’EXPORT SETTORIALE nel 2011-2012

Nel biennio preso in considerazione nella relazione annuale ISTAT, particolare enfasi è posta sulla rilevanza di un dato: il crollo della domanda interna nel nostro Paese ha fornito notevole slancio verso l’export, migliorando di molto la Bilancia Commerciale.

Il crollo dei consumi genera un ribasso consistente delle importazioni, e come in un circolo virtuoso, dato che il consumo interno si assesta su livelli più bassi, tutto quanto eccede viene ad essere esportato.

E' proprio questo il vero motivo per cui nel biennio 2011-2012 le esportazioni italiane hanno continuato a segnare saggi di crescita rilevanti.

Rileviamo inoltre che tra le imprese distintesi all'estero è presente un segmento di aziende cosiddette "globali", le quali a differenza delle semplici imprese esportatrici, sono di gran lunga più internazionalizzate, ed il cui impatto in termini di occupazione, PIL e valore aggiunto è stato enorme: il 7% in più nell'ambito occupazionale, e l'8% sul valore aggiunto.

Spostando l'analisi dalle imprese "globali" (quelle esportatrici internazionalizzate) a quelle multinazionali, otterremo un effetto ancor più amplificato in termini percentuali (+9% di occupazione e +13% in termini di valore aggiunto).

Se addizioniamo a questo ulteriori elementi incoraggianti riguardo alla congiuntura internazionale prevista per il prossimo biennio, che dovrebbe concretizzarsi in un incremento della domanda mondiale per i prossimi anni, c'è da essere ottimisti circa i trend di sviluppo dell'economia italiana.

Per quel che concerne invece l'andamento dell'export italiano, nel 2012 la situazione è mutata radicalmente: dalla tabella seguente si evince la chiara catalogazione dei prodotti più rilevanti in termini di valore aggiunto. Essi sono nove in tutto, e si accaparrano una "fetta" molto rilevante nella "torta"

del valore aggiunto italiano proveniente dalle esportazioni.

Tavola 2.5 Analisi della dinamica delle esportazioni settoriali – Anni 2010-2012 (valori percentuali)

DIVISIONI ATECO	Contributi percentuali alla variazione totale del 2012 rispetto al 2010	Quote in valore	Variazioni percentuali rispetto all'anno precedente			Contributi percentuali alla variazione totale			Quote di esportazione nell'area Ue			Variazioni percentuali degli indici dei volumi esportati verso l'area Ue		Variazioni percentuali degli indici dei volumi esportati verso l'area extra-Ue	
			2012	2011	2012	2011	2012	2010	2011	2012	2011	2012	2011	2012	
Prodotti delle attività manifatturiere	96,5	95,8	11,6	3,6	97,0	95,0	56,9	55,7	53,4	1,4	-3,6	8,0	3,1		
DIVISIONI ATECO A RILEVANTE CONTRIBUTO POSITIVO ALLA VARIAZIONE DELLE ESPORTAZIONI															
Macchinari e apparecchiature n.c.a.	19,9	18,1	14,0	3,0	21,8	14,7	43,5	42,9	41,3	6,2	-5,7	11,4	-0,5		
Prodotti della metallurgia	18,2	8,4	32,4	6,4	19,6	14,3	62,7	60,6	55,7	14,7	2,8	25,2	15,0		
Coke e prodotti derivanti dalla raffinazione del petrolio	10,9	5,3	13,9	21,8	5,3	26,5	42,2	38,3	34,2	-20,7	-4,7	-7,0	11,7		
Prodotti farmaceutici di base e preparati farmaceutici	6,2	4,4	9,6	12,5	3,5	13,8	56,8	55,1	58,5	-1,9	11,5	10,6	-0,8		
Articoli in pelle (escluso abbigliamento) e simili	6,0	4,2	16,6	5,9	5,8	6,6	52,7	50,4	47,9	3,0	-5,0	13,3	2,8		
Prodotti alimentari	5,5	5,1	9,8	6,6	4,3	8,8	69,7	68,6	66,9	1,4	0,3	8,9	9,0		
Prodotti chimici	5,3	6,5	10,4	1,6	6,1	2,9	61,9	62,5	61,0	0,9	-3,9	1,6	0,3		
Autoveicoli, rimorchi e semirimorchi	4,5	6,4	11,1	-0,6	6,5	-1,1	69,8	70,0	66,5	8,6	-6,5	11,8	6,9		
Articoli di abbigliamento (anche in pelle e in pelliccia)	4,1	4,4	10,8	3,2	4,2	3,8	56,5	53,6	50,8	-1,5	-7,3	10,6	0,8		
DIVISIONI ATECO CHE FORNISCONO BASSO CONTRIBUTO ALLA VARIAZIONE DELLE ESPORTAZIONI															
Altri mezzi di trasporto	-1,4	2,9	-4,2	-1,9	-1,3	-1,6	52,9	41,8	42,3	-30,9	-1,3	11,4	-4,6		
Tabacco	0,0	0,0	18,7	21,2	0,0	0,0	57,6	45,7	37,5	-6,7	-4,6	52,1	24,3		
Prodotti della stampa e della riproduzione di supporti registrati	0,0	0,0	-10,5	35,3	-0,0	0,1	66,8	68,3	75,2	-12,2	42,9	-20,6	-10,1		
Legno e prodotti in legno e sughero (esclusi i mobili);	0,2	0,4	3,4	4,4	0,1	0,5	57,0	56,6	52,9	-0,3	-3,6	2,1	9,6		
Articoli in paglia e materiali da intreccio	0,7	1,6	5,4	0,8	0,8	0,3	70,8	70,6	69,3	0,6	-0,3	1,5	3,7		
Carta e di prodotti di carta	0,7	2,1	3,9	0,9	0,8	0,5	60,8	59,1	55,8	-5,7	-6,5	6,9	4,5		
Altri prodotti della lavorazione di minerali non metalliferi	0,8	2,3	2,8	2,2	0,6	1,4	56,8	57,8	54,9	-0,9	-6,2	-0,9	4,0		
Prodotti tessili	0,9	2,4	9,0	-3,6	2,1	-2,5	59,4	59,4	57,7	0,3	-10,1	-0,3	-5,9		
Apparecchiature elettriche e apparecchiature per uso domestico non elettriche	1,1	5,1	4,8	-1,8	2,4	-2,7	61,9	60,9	59,8	-1,9	-5,2	8,8	2,2		
Totale prodotti	100,0	100,0	11,4	3,7	100,0	100,0	57,3	56,0	53,7	1,1	-3,7	7,9	3,2		

Fonte: Elaborazioni su dati Istat

(TABELLA FONTE ISTAT).

Un'ulteriore tendenza positiva è rappresentata dalla continua espansione dei prodotti italiani al di fuori del "recinto" UE; a monte possono essere considerate come rilevanti due tipi di ragioni: in primis battuta l'annosa questione circa la depressione della domanda di consumi interna. Ma questa da sola non basta: se considerassimo solo il crollo della domanda interna, spiegheremmo, infatti, semplicemente l'aumento delle esportazioni, e non l'incremento relativo dell'export extra-europeo.

Perciò consideriamo in secundis, sia la capacità delle imprese italiane di "riposizionarsi" all'interno del panorama di commercio globale, sia l'inerzia, nell'accezione positiva del suo significato, che ha portato semplicemente al consolidamento di un trend già presente in passato.

L'ultima componente, e non per importanza, che ha inciso sempre positivamente sull'andamento del commercio extraeuropeo, è stata l'efficienza di costo manifestatasi nei confronti della gestione del commercio extra continentale (questo ultimo punto non presenta però differenziali in termini percentuali significativamente diversi da zero, ragion per cui, consideriamo solo i primi due elementi come le variabili più incisive al miglioramento dell'export extra-europeo).

ANALISI DELLE VARIABILI DEL MODELLO DI REGRESSIONE:

Focalizzando l'attenzione sulle variabili che considereremo nell'analisi di regressione, in cui dimostreremo statisticamente l'incidenza dello scambio di merci col mondo sul pil pro capite, possiamo applicare una distinzione preliminare tra le tipologie di variabili che andremo ad inserire nella stima del modello, suddividendole in più macroaree, a seconda della natura del loro impatto sulla nostra variabile dipendente (Pil pro-capite).

Volendo dividere le stesse variabili per fasce, partiamo da quella di "crescita, struttura e performance", passando per quella relativa all'"internazionalizzazione", categoria su cui presteremo particolare interesse; in ultima analisi considereremo i "fattori strutturali di contesto".

Questa tripartizione è di cruciale importanza all'interno dell'analisi di regressione che sarà svolta in seguito, dato che le variabili del modello che considereremo nel capitolo 3, vengono tratte da ognuno di questi tre gruppi.

Nella fattispecie, considereremo le seguenti variabili: il PIL Pro-Capite, il Costo del Lavoro, il valore aggiunto delle grandi imprese, il numero di imprese estere a partecipazione italiana, gli scambi di merci col mondo, le imprese a partecipazione estera e la spesa in ricerca e sviluppo.

Tra le variabili di “crescita struttura e performance”, annoveriamo sicuramente il costo del lavoro, il quale impatta sulla produttività del business, ed il valore aggiunto delle grandi imprese: esso incide notevolmente anche sul commercio internazionale stesso. Sono le grandi imprese infatti, quelle più internazionalizzate, in alcuni casi addirittura “multinazionali”, ad incidere decisamente per il loro livello di volumi di vendita sull’intero ammontare del prodotto interno esportato.

Per quel che concerne invece le variabili di “internazionalizzazione”, consideriamo lo scambio di merci col mondo, in misura percentuale del pil, come valore “proxi” delle esportazioni al netto delle importazioni.

Proseguendo, passiamo al numero di imprese estere partecipate e le imprese italiane a partecipazione estera. Queste due variabili ci danno un chiaro indice del grado di apertura della nostra economia a livello di investimenti, e non di semplice attività commerciale.

In ultima battuta, troviamo la spesa in Ricerca e Sviluppo, tra i “fattori strutturali di contesto”, quello senza dubbio più rilevante ai fini della comprensione della reiterata competitività all’estero dei nostri prodotti.

Terminata la breve introduzione, in cui abbiamo analizzato sommariamente la situazione economica generale e del settore industriale italiano, passiamo dunque alla spiegazione dei preliminari statistici basilari alla spiegazione del modello di regressione lineare multivariato che verrà illustrato successivamente.

CAPITOLO 2

All'interno di questo capitolo, verranno esposti i fondamenti indispensabili all'analisi di regressione multipla necessari alla comprensione delle argomentazioni che saranno esposte nel capitolo 3.

IL MODELLO DI REGRESSIONE:

Innanzitutto procediamo con la definizione di modello di regressione. Esso è principalmente un *“modo di rappresentare la realtà, mediante il quale descrivere ed interpretare fenomeni di interesse”*. (Anna Clara Monti, “Introduzione alla Statistica”)

Ma oltre alla mera previsione o interpretazione, esso fornisce inoltre la capacità di prevedere quelli che saranno gli andamenti futuri di una certa variabile, o di un gruppo di variabili.

Nella fattispecie, possiamo distinguere e classificare i modelli di regressione in base al numero di variabili che contengono, alla forma funzionale su cui si basano, ecc..

LE VARIABILI: DIPENDENTI ED INDIPENDENTI

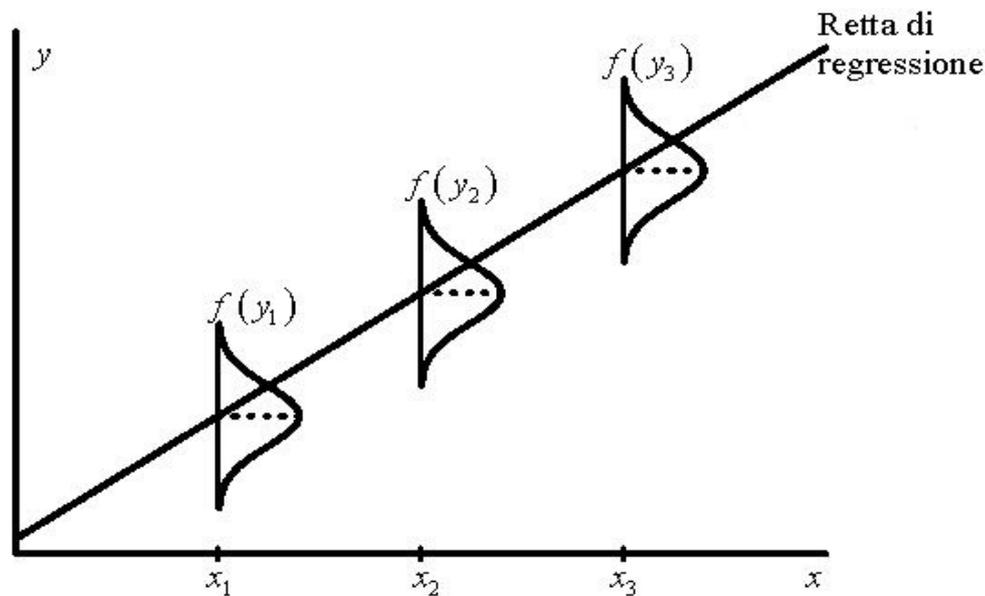
La classificazione più importante riguarda la natura delle variabili. Distinguiamo infatti le variabili del modello di regressione in variabile “dipendente” e variabile “indipendente” o “regressore”.

Primo assunto fondamentale è il legame presente tra le due variabili, il quale in media, ossia in termini di valore atteso, fa sì che il “*valore atteso della Y sia funzione del valore assunto dalla X*” (Anna Clara Monti, Introduzione alla Statistica). Il suddetto legame può attuarsi, a seconda dei dati che sono a disposizione, in una forma funzionale lineare (retta) o non lineare (iperbole, logaritmo...). Nel nostro caso, ci occuperemo di regressione lineare, in cui il valore atteso di Y legato al regressore, si troverà in media esattamente sulla retta (funzione lineare) di regressione.

IL MODELLO DI REGRESSIONE COME ESEMPLIFICAZIONE DELLA REALTÀ

Il modello di regressione è un'esemplificazione della realtà: ci dice infatti, quelli che in media saranno i valori predetti al variare del regressore sulla variabile dipendente. Ma dato che si tratta di esemplificazione della realtà, non tutte le osservazioni saranno “spiegate dal modello”, e cioè alcune non si troveranno sulla retta stessa, ma in un intorno più o meno vasto della linea di tendenza predetta.

Questa che potrebbe sembrare a prima vista una mancanza, non ne limita l'efficacia e l'utilità: infatti, pur considerando che le osservazioni dei dati empirici non si andranno a posare precisamente sul valore predetto (ossia sulla retta), esse potranno essere reperite per l'appunto, con un certo livello di confidenza, nel suddetto intorno esteso o ridotto, a seconda della variabilità del modello.



(Fonte immagine, Anna Clara Monti: "Introduzione alla statistica")

la seguente retta di regressione: $Y_i = \beta_0 + \beta_1 X_1 + u_i$

ci spiega come per l' i-esimo dato (ma in realtà il discorso è esteso ad ogni osservazione) il valore osservato, ossia il "valore vero", con probabilità pari al valore atteso della gaussiana costruita su ogni osservazione, sarà nello stesso punto del valore predetto, ossia sulla retta di regressione stessa.

Dato quindi che il valore atteso per ogni osservazione si trova sulla retta stessa, allora proprio lungo tutta la retta ci sarà una più elevata probabilità di matching tra valori "predetti" e valori "veri".

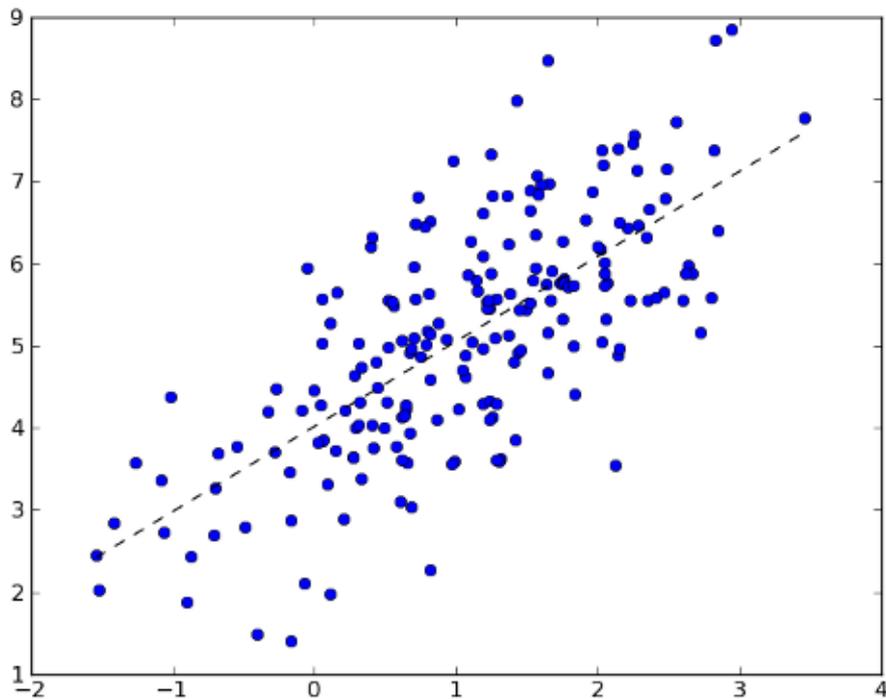
Per quanto concerne la regressione nel suo complesso, la Y rappresenta la variabile dipendente, il cui andamento è condizionato dal regressore X, secondo la proporzione indicata dal coefficiente angolare β_1 . L'intercetta, ossia il punto in cui è soddisfatta la condizione di "X=0" è rappresentata da β_0 , mentre u_i è la componente stocastica del modello. La componente stocastica, o errore, è quanto non viene ad essere catturato dal modello

“deterministico”, e che residua ad esso: il modello deterministico, con le sue componenti (β_1 , β_0 e le variabili in correlazione), pur riuscendo a spiegare la relazione causale tra X ed Y, non riesce a spiegare a pieno la correlazione.

Il modello di regressione, pertanto oltre ad avere una componente deterministica, ne possiede una stocastica, la cui presenza genera delle implicazioni: una componente stocastica all'interno di un modello di regressione implica però variabilità, per la sua natura di aleatoria di variabile casuale.

Ed è proprio l'errore stocastico che va ad inficiare sulla bontà del modello di regressione costruito: tanto più esso è piccolo, quanto la gaussiana costruita su ogni punto (nell'immagine precedente) risulta essere più schiacciata; per cui, se la varianza della gaussiana di un punto sarà bassa, con molta probabilità l'osservazione o cadrà precisamente sulla retta, nel caso in cui l'errore stocastico è pari a zero, oppure in un suo intorno molto piccolo.

Infatti, proprio sulla presenza dell'errore residuo gli studiosi calcolano la retta di regressione che meglio di tutte si applica al modello: la tecnica utilizzata per ottenere la retta di regressione, è quella cosiddetta OLS (Ordinary Least Square), cui viene individuata la retta di regressione che soddisfi la condizione di minimizzare la distanza di tutti i residui, o errori.



Nell'immagine riportata, notiamo come la retta di regressione, minimizzando la somma dei residui, elevati al quadrato, si posi esattamente nei punti in cui in media l'errore stocastico è il più basso di tutti. Dato che i residui possono essere sia negativi che positivi, onde evitare problemi di calcolo e di riporto dei segni, essi vengono elevati al quadrato in maniera tale che la retta non subisca distorsioni nel senso positivo o negativo a seconda dei casi.

IL METODO DEI MINIMI QUADRATI

Per stimare i coefficienti del modello di regressione, di cui ci occuperemo nel prossimo capitolo, ci serviremo dei dati a campione rilevati dall'ISTAT circa le variabili che andremo a considerare come rilevanti statisticamente nel

loro effetto sul pil procapite. Il nostro modello, diversamente da quelli semplici (i quali contengono solo una variabile indipendente), essendo multivariato, catturerà la presenza di più di una variabile indipendente .

Passo cruciale nella stima del modello, è l'individuazione di una giusta intercetta, e di adeguati coefficienti da moltiplicare a ciascun regressore; ne consegue dunque, che risulterà di fondamentale importanza riuscire a capire qual'è la retta che meglio di tutte interpola la "nuvola di punti" rilevati a campione.

Come già anticipato, il modello di cui ci serviremo sfrutterà più variabili come regressori, i cui coefficienti stimati (indicati con "b") saranno quindi quelli per cui sarà soddisfatta la seguente condizione:

$$RSS = \sum \hat{u}_i^2 = \sum (Y - b - b_1 X_{i1} - b_2 X_{i2} - b_3 X_{i3})^2 = \text{MIN}$$

$$\frac{\partial RSS}{\partial b_1} = 0 ; \frac{\partial RSS}{\partial b_2} = 0 ; \frac{\partial RSS}{\partial b_3} = 0$$

... E LE SUE IPOTESI:

Laddove venga appurata l'esistenza delle condizioni spiegate nel precedente paragrafo, possiamo definire delle ipotesi che l'OLS richiede affinché sia efficace:

Assunto 1: Il modello è lineare nei parametri e correttamente specificato:

Assunto 2: Il valore atteso del termine di disturbo è pari a zero.

$E(U_i) = 0$ per ogni i

Assunto 3: Il termine di disturbo è omoschedastico: ossia la varianza dell'errore stocastico è u considerata uniforme, quindi costante.

Assunto 4: I termini di disturbo sono distribuiti indipendentemente → la covarianza di U_i e U_j è pari a zero, per ogni i -esimo e j -esimo elemento

Assunto 5: I termini di disturbo si distribuiscono normalmente.

Ma dato che lo stimatore OLS è di per sé una variabile casuale, in quanto dipende da Y , e transitivamente anche da u , allora possiamo affermare che \mathbf{b}_0 e \mathbf{b}_1 saranno a loro volta caratterizzati da una loro distribuzione campionaria, che inficerà in primis sulla correttezza (1) dello stimatore, in secundis sulla sua variabilità (2).

L'OLS è stato dimostrato essere lo stimatore più efficiente tra quelli non distorti: ossia, dati tutti gli stimatori corretti (1), è quello con varianza minore, all'interno del quale le osservazioni sono più concentrate nell'intorno minimo (relativo agli altri stimatori) della media (2).

Ed è proprio in virtù di questa peculiarità che si è guadagnato l'acronimo di stimatore "BLUE", ossia "Best Linear Unbiased Estimator", ossia il miglior stimatore in termini di efficienza (varianza bassa) tra i non distorti.

Per quanto riguarda la varianza, è necessario specificare altro: essa infatti l'indice più importante di "efficienza" dello stimatore, ossia, di capacità di posizionare con elevata confidenza (quindi sulla base di una probabilità molto buona) i valori del modello quanto più è vicino possibile al valore vero.

Questo però non è sempre possibile, in quanto a sua volta, la varianza stessa è funzione sia del numero di osservazioni, sia della varianza campionaria delle X .

"Quanto più elevata sarà la dispersione delle esplicative tanto maggiore è l'informazione e quindi la precisione delle stime del parametro" (Stock James H.; Watson Mark W., "Introduzione all'Econometria"). Infatti, più il modello risulterà impreciso, meno avremo confidenza nei parametri stimati. Così come, più elevato sarà la numerosità campionaria tanto più avremo informazioni utili per stimare il modello.

MODELLO DI REGRESSIONE MULTIVARIATO:

(riferimenti S.Borra, A. Di Ciaccio – Statistica, metodologie per le scienze economiche e sociali, Mc Graw Hill,2004)

A questo punto è opportuno spendere qualche parola in più sul modello di regressione multivariato.

Possiamo considerare il modello di regressione multivariato come “un’ estensione della regressione lineare semplice al caso in cui si impegnano due o più variabili esplicative per dar conto del comportamento della variabile risposta”.

Data la notevole importanza dell’argomento, ed il contenuto spazio entro cui lo tratteremo, definiamo i tre punti cruciali che contraddistinguono il modello di regressione lineare multipla e le sue principali peculiarità: la sua forma matriciale, la stima puntuale dei parametri, ad in ultimo l’inferenza su di essi mediante test d’ipotesi.

La forma matriciale:

L’algebra matriciale applicata al modello di regressione multivariato, ci consente di sintetizzare molto più agevolmente i risultati del modello. Consideriamo Y , β ed u , ma questa volta non come singoli numeri, bensì come vettori: rispettivamente Y sarà il vettore colonna della variabile dipendente, β sarà il vettore dei coefficienti, ed u quello dei residui.

$$Y = \begin{matrix} y_1 \\ y_2 \\ \dots \\ y_n \end{matrix} \quad \beta = \begin{matrix} \beta_0 \\ \beta_2 \\ \dots \\ \beta_k \end{matrix} \quad u = \begin{matrix} u_1 \\ u_2 \\ \dots \\ u_n \end{matrix}$$

E sia X la matrice:

$$X = \begin{matrix} & 1 & X_{11} & \dots & X_{1k} \\ = & 1 & X_{21} & \dots & X_{2k} \\ & \vdots & \vdots & \ddots & \vdots \\ & 1 & X_{n1} & \dots & X_{nk} \end{matrix}$$

Dove la prima colonna rappresenta l'intercetta, dalla seconda in poi sono riportati i valori delle esplicative.

In presenza dei seguenti tre assunti:

1) $Y = X\beta + u$

2) u è un vettore il cui valore atteso $E(u)=0$, quindi è stocastico ed indipendente.

3) X è una matrice di costanti ed ha rango pieno.

Possiamo considerare che: X è una matrice di dimensione $n \times (k + 1)$ e giacchè β è un vettore (colonna) di dimensione $(k+1) \times 1$, ne consegue che $X\beta$ sarà un vettore di dimensione $n \times 1$, il cui i-esimo elemento sarà pari a:

$$" \beta_0 + \sum_{j=1}^k \beta_j X_{ij} "$$

In più, dato che $E(u) = 0, \forall 1 < i < n$, otteniamo la "Matrice di Varianza e Covarianza" equivalente a:

$$V(u) = \sigma^2 I = \sigma^2 \begin{matrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{matrix} = \begin{matrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{matrix} \quad \text{I termini d'errore sono}$$

incorrelati, in quanto $COV(u_i, u_j) = 0 \forall i = 1, 2, \dots, n$ con $j \neq i$ e laddove

$$V(u_j) = \sigma^2 \forall i = 1, 2, \dots, n .$$

Il vettore delle Y ha un valore atteso pari ad $X\beta$, ed una matrice di varianze e covarianze pari a $\sigma^2 I$.

Per quel che riguarda [la stima puntuale dei parametri](#), è proprio in essa che riscontriamo il maggiore vantaggio di essere ricorsi all'algebra lineare: essa ci consente di ricondurre alla seguente equazione la stima del vettore dei β :

$$b = (X'X)^{-1} X' Y \quad \text{(consideriamo "b" come stimatore corretto per il modello OLS multivariato)}$$

Così facendo, evitiamo per ogni coefficiente di porre la condizione di minimizzare gli scarti e di ricalcolarci il valore di β come per il modello lineare semplice.

Avendo infatti il vettore già sotto questa forma, siamo in grado di stabilire quelli che sono i coefficienti di ogni singola variabile indipendente, con un evidente vantaggio in termini di praticità.

[Inferenza sui parametri del modello di regressione:](#)

Per poter inferire circa gli stimatori del modello di regressione multivariato, sono opportune delle osservazioni preliminari: nel modello di regressione multivariato, infatti lo stimatore OLS di β ha distribuzione normale (nel caso della regressione multipla consideriamo la distribuzione normale multivariata), con:

- 1) vettore dei valori attesi pari a $E(b)=\beta$ (stimatore corretto)
- 2) matrice di varianza e covarianza $V(b) = \sigma^2 (X'X)^{-1}$

Per cui, avremo che il vettore b si distribuisce, in ogni sua componente, come una normale multivariata con valore atteso del vettore pari ad 1) e matrice di varianza e covarianza pari a 2) :

$$b \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Da ciò discende che ogni elemento del vettore b avrà distribuzione Normale, in cui il

j -esimo si distribuisce così:

$$b_j \sim N(\beta_j, \sigma^2 c_{jj}), \text{ per } j=0,1,\dots,k.$$

Otteniamo quindi una statistica test pari a :

$$\frac{b_j - \beta_j}{\sigma \sqrt{c_{jj}}} \sim N(0,1).$$

Ma dato che nella realtà è impossibile conoscere la σ al denominatore, si ricorre allo stimatore S^2 , il quale sostituito alla statistica test ci porta al seguente risultato:

$$\frac{b_j - \beta_j}{S(\beta_j)} \sim t_{n-k-1}$$

La distribuzione studentizzata discende dal fatto che, senza addentrarci troppo nel merito, il rapporto tra una normale standard (al numeratore) ed una χ^2 al denominatore, si distribuisce a sua volta come una t di Student.

Da questo rapporto ricaviamo gli intervalli di confidenza per i test d'ipotesi sui coefficienti:

$$IC: [b_j \pm t^{\alpha/2} s(b_j)]$$

I test d'ipotesi che via via andremo ad affrontare nel capitolo successivo, saranno per lo più basati su questo tipo di statistica test: andremo a testare cioè la significatività di ogni singolo coefficiente (ma il principio è lo stesso che per la regressione congiunta) laddove ne dimostreremo la loro diversità da zero.

Il sistema d'ipotesi che considereremo sarà del tipo:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Con una regione d'accettazione dell'ipotesi nulla pari a:

$$H_0: |t| < t^{\alpha/2}$$

Per la significatività congiunta dell'intera regressione, rimandiamo al prossimo capitolo in cui analizzeremo l'output della regressione applicando anche la "statistica F", ossia proprio il test che è utile ad inferire circa la significatività congiunta del modello.

IL COEFFICIENTE DI DETERMINAZIONE

Dato il nostro obiettivo di spiegare il modello per cui le esportazioni italiane impattano sul PIL pro-capite, un altro elemento di cui ci serviremo nel prossimo capitolo, sarà l'indice di determinazione, più comunemente chiamato "R²".

Esso ci consente di capire la percentuale di "variabilità della Y spiegata dal modello di regressione".

Essendo un numero puro, R^2 varia da zero a uno e quanto più è prossimo al valore unitario, tanto più il modello stimato avrà una maggiore bontà di adattamento alla realtà: laddove raggiungesse il valore uno, starebbe a significare che il modello di regressione in quel caso catturerebbe al 100% tutta la variabilità della sua Y , e quindi non avremmo alcuna differenza tra modello vero e modello predetto.

Dopo aver richiamato alla mente le nozioni principali e propedeutiche ad un'analisi di regressione, possiamo procedere alla stima vera e propria del modello di regressione.

CAPITOLO 3

Nell'ultimo capitolo di questo lavoro, verrà presentato il frutto della ricerca statistica basata sui dati forniti dall' ISTAT, volta ad interpretare gli indici di sviluppo a livello regionale e nazionale dei fattori che hanno contribuito in misura significativa nel loro apporto al prodotto nazionale.

Al termine della ricerca, saremo in grado di analizzare statisticamente quelle che sono conseguenze di cui, qualitativamente si sente parlare ogni giorno, ma che in termini numerici risulta complesso commentare.

Ci serviremo inoltre di Microsoft Office Excel per costruire la retta di regressione, e per analizzare quelli che sono i possibili scenari futuri.

L'analisi che effettueremo prevederà in via preliminare la selezione delle variabili del modello mediante metodo "Best Subset", che imposteremo grazie alla macro "PhStat", la quale ci fornirà i risultati per tutte le combinazioni possibili delle variabili indipendenti.

Una volta individuato il modello migliore in base al valore di R^2 corretto(per il numero di variabili), passeremo all'analisi dei coefficienti della regressione, ed in fondo forniremo tutti i commenti relativi all'output che conseguirà allo studio del modello.

Dapprima, presentiamo il campione che utilizzeremo per la ricerca, riportato dall'ISTAT.

Y	X1	X2	X2
PIEMONTE	28276,87632	0,470973246	112496,199
VALLE D'AOSTA	35264,84699	0,395300913	3929,032839
LOMBARDIA	33483,8088	0,462969971	300901,8969
TRENT. A. ADIGE	33556,28592	0,460046455	31602,32742
VENETO	29881,89117	0,473014495	133624,4085
FRIULI V. GIU.	29401,7353	0,500494982	32942,54569
LIGURIA	27396,17919	0,441524701	39323,42367
ECMILIA ROM.	31688,87542	0,487045347	126072,3112
TOSCANA	28209,39031	0,451425195	94404,14222
UMBRIA	23988,86967	0,455751075	19529,99255
MARCHE	26412,16118	0,457721291	36985,12751
LAZIO	29430,00002	0,487185827	153165,0491
ABRUZZO	22061,97077	0,484240841	26336,80111
MOLISE	20173,12475	0,426638096	5771,13067
CAMPANIA	16601,24132	0,495394379	85038,32536
PUGLIA	17545,53031	0,485487241	63402,20093
BASILICATA	18437,38503	0,462809335	9747,232492
CALABRIA	16876,55006	0,452979009	29753,85576
SICILIA	17189,0014	0,478872121	76128,71616
SARDEGNA	20071,37218	0,486574763	29962,01263
X3	X4	X5	X6
0,104124923	0,0568	1,841213907	2.806,00
0,073860438	0,036298	0,655989499	15,00
-0,015593209	0,110296	1,271754878	9.647,00
0,026331226	0,0389595	1,306104465	554,00
0,090114799	0,026821	1,079062288	3.679,00

0,127795368	0,036275	1,485049922	759,00
-0,094815455	0,046343	1,358186317	528,00
0,150023203	0,034618	1,371822025	3.554,00
0,095044348	0,026029	1,226190225	1.443,00
0,065472128	0,020564	0,979053753	187,00
0,076004402	0,00925	0,705572216	685,00
-0,068666423	0,07565	1,818727493	2.257,00
0,116083649	0,050769	0,941320656	195,00
-0,002637526	0,003412	0,502499887	37,00
-0,01636536	0,006991	1,259965432	392,00
0,039659786	0,006761	0,767917941	183,00
0,039659786	0,00159	0,672509856	30,00
-0,00613433	0,003479	0,459487586	22,00
-0,091556357	0,002922	0,855829968	144,00
-0,126319488	0,006341	0,665608842	74,00

Le variabili presenti nella tabella sono rispettivamente :

Y) Pil pro-capite (variabile dipendente)

X1) COSTO DEL LAVORO (h/b %)

X2) valore aggiunto (milioni di €)

X3) Scambio di merci con il mondo (come % del PIL)

X4) addetti imprese a partecipazione estera.

X5) spesa in Ricerca&Sviluppo (come % del PIL)

X6) addetti imprese estere partecipate .

A partire dalla precedente tabella, conduciamo mediante la macro di Microsoft Office Excel chiamata "PhStat" l'analisi che ci consentirà di scoprire quali di queste variabili meglio interpretano nonchè catturano l'effetto dello "Scambio di Merci don il Mondo" nei confronti del Pil pro-capite.

ANALISI DI REGRESSIONE: DUE CASI ESTREMI

Volendo procedere per gradi, consideriamo dapprima il caso più semplice di analisi di regressione, quello a due variabili, in cui la Y è rappresentata dal pil procapite, e la X, è la "X3" precedentemente introdotta, ossia lo scambio di merci col mondo. Dopodichè, ci focalizzeremo sul caso opposto in cui considereremo tutte le variabili della tabella istat (casi estremi del modello).

Nel primo caso, possiamo anticipare che tale regressione lineare (semplice) risulterebbe poco utile, come viene palesato dall'output di Excel:

<i>Regression Statistics</i>	
Multiple R	0,3538
R Square	0,1252
Adjusted R Square	0,0737
	6125,818
Standard Error	9
Observations	20

Calculations

<i>b1, b0 Coefficients</i>	28123,2640	24433,0735
<i>b1, b0 Standard Error</i>	18032,5503	1476,7563
R Square, Standard Error	0,1252	6125,8189
<i>F, Residual df</i>	2,4323	17,0000
Regression SS, Residual SS	91273656,076	637936166,985
	6	4

Confidence level	95%
<i>t</i> Critical Value	2,1098
Half Width <i>b0</i>	3115,6835
Half Width <i>b1</i>	38045,3555

Come si evince dall'output della regressione, il modello risulta essere abbastanza blando nella sua rilevanza. Esso presenta un coefficiente R square bassissimo: quindi la percentuale di devianza della variabile dipendente catturata dal modello risulta essere bassa.

In più, analizzando nello specifico:

<i>P-</i>				
<i>value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
0,0000	21317,3900	27548,7570	21317,3900	27548,75704
0,1373	-9922,0915	66168,6195	-9922,0915	66168,61945

I coefficienti *b0* e *b1* risultano essere, il primo significativo, mentre il secondo non significativo: il che implica come testando l'ipotesi nulla in cui la pendenza della retta di regressione sia uguale a zero, verso quella alternativa in cui la pendenza risulti diversa da zero, con un *p* - value del

13,73% non potremmo rifiutare l'ipotesi nulla nemmeno con una confidenza pari al 90%.

Ergo, il modello non risulta statisticamente rilevante.

Una volta appurata la scarsa utilità del modello di regressione lineare semplice a sole due variabili, in cui la retta di regressione non ha coefficienti significativi, ed in cui l'R square è bassissimo, passiamo considerare l'ipotesi diametralmente opposta in cui includiamo tutte e sette le variabili presenti all'interno della tabella ISTAT. L'output che rileveremmo sotto quest'altra ipotesi è il seguente:

Regression Analysis

Regression Statistics				
Multiple R	0,9041			
R Square	0,8174			
Adjusted R Square	0,7261			
Standard Error	3331,3444			
Observations	19			
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	66789,7851	17189,5391	3,8855	0,0022
-				
0,470973246	113736,1557	39958,6397	-2,8463	0,0147
112496,199	-0,0658	0,0441	-1,4920	0,1615
0,104124923	15821,0509	11562,4414	1,3683	0,1963
0,0568	29441,1171	59402,9858	0,4956	0,6291
1,841213907	11204,5485	3765,2862	2,9757	0,0116

2806	2,4361	1,4193	1,7165	0,1118
------	--------	--------	--------	--------

In questo caso, l'R square risulta essere molto più elevato, in quanto risente positivamente dell'inclusione di variabili rilevanti, che nel modello precedente ad una singola variabile indipendente erano state omesse.

A fronte di un miglior adattamento del modello ai dati però, riscontriamo irrilevanza statistica per ben quattro variabili, i cui p-value sono >10%.

Il p-valore, ossia il valore soglia usato come punto di riferimento per testare la differenza "statistica" di una variabile da zero, ci consente di affibbiare ai coefficienti delle variabili l'appellativo di "statisticamente rilevante" a seconda che esso superi o meno un valore soglia.

Dato che nel modello a sette regressori il p-valore risulta più grande del 10% in molti casi, non ci dà con buona probabilità la sicurezza relativa che i coefficienti il cui p-valore è maggiore del livello soglia, siano effettivamente rilevanti: il che ci induce a pensare che tale differenza, in effetti, non ci sia, e che quindi le variabili stesse risultino uguali a zero (Statisticamente Irrilevanti).

Questo modello quindi, per le motivazioni supportate nel precedente paragrafo non può essere applicato alla nostra ricerca.

MODELLO DI REGRESSIONE SELEZIONATO MEDIANTE METODO BEST SUBSET:

Il metodo "best subset", come precedentemente accennato, ci consente di individuare un modello che risponda al meglio alla nostra esigenza di

testare la significatività del livello su cui si attesta lo scambio di merci col mondo, rispetto al PIL Pro-capite.

Dato l'elevatissimo numero possibile di combinazioni di modelli lineari multivariati che potrebbero essere calcolati sulla base di sette variabili indipendenti, ci serviamo della regressione mediante "best subset": questa tecnica, prevede una parte meramente meccanica, ossia quella in cui il software calcola ogni possibile combinazione tra le variabili.

La fase inferenziale successiva, consente di individuare il modello "migliore" (in termini di adattamento ,ossia di R^2) tra tutti quelli che vengono forniti dal software stesso.

Utilizzando quindi questa tecnica, siamo giunti alla conclusione che il modello in cui vengono considerate le variabili "Costo del Lavoro", "Spesa in R&S % del PIL", e gli "Scambi con il mondo (%PIL)" è quello che meglio di tutti si adatta allo scopo della nostra ricerca.

Questo modello, lineare nei parametri e multivariato, ci consente di arrivare al migliore risultato possibile statisticamente degno di nota.

IL MODELLO :

Procediamo illustrando l'output del software che otteniamo dalla regressione delle tre variabili "Costo del Lavoro", "Spesa in R&S % del PIL", e gli "Scambi con il mondo (%PIL)" sul Pil Procapite. Il risultato è il seguente:

<i>Regression Statistics</i>	
Multiple R	0,8348
R Square	0,6969
Adj R Square	0,6363
Standard Error	3838,6059
Observations	19

	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>
Intercetta	80284,9808	171444,6663	4,6828
-	-	-	-
Costo del lavoro	148388,3566	38975,6066	-3,8072
Sc. Merci (%PIL)	21536,73185	11760,63886	1,831255267
Spesa R&S (% PIL)	13112,2561	2714,1758	4,8310

Con una significatività ed un intervallo di confidenza rispettivamente di:

<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
<i>0,0003</i>	<i>43741,9896</i>	<i>116827,9719</i>
<i>0,0017</i>	<i>-231462,8957</i>	<i>-65313,8176</i>
<i>0,0434</i>	<i>839,6488</i>	<i>49084,7328</i>
<i>0,0002</i>	<i>7327,1273</i>	<i>18897,3849</i>

Il precedente output, a differenza dei due casi estremi prima illustrati, mostra subito a primo impatto che ognuna delle sue variabili risulta statisticamente significativa.

Inoltre, ci consente di affermare con relativa tranquillità che in termini di coefficiente di correlazione ci sia una buona parte di variabilità catturata dal modello in termini di bontà di adattamento ($R^2= 63\%$).

In più, la significatività delle variabili non è minata e non supera mai la soglia $\alpha=5\%$ in nessun parametro considerato.

Per quanto riguarda le variabili “Costo del Lavoro” e “Spesa in R&S % del PIL”, il p-value è bassissimo, per cui possiamo con un’elevatissima probabilità inferire circa la significativa diversità da zero dei coefficienti che premoltiplicano le suddette variabili.

Per quanto riguarda gli “Scambi con il mondo (%PIL)”, con un p-value del 4,3%% circa, e quindi con una confidenza di più del 95%, perveniamo alla stessa conclusione.

La significatività singola dei tre coefficienti, requisito fondamentale per considerare “buono” un modello, si unisce ad una chiara e netta significatività congiunta del modello nel suo insieme.

<i>F</i>	<i>Significatività F</i>
11,4962	0,0004

La statistica test “F”, ci rivela quindi che i coefficienti del nostro modello con una confidenza del 99% risultano essere significativi se considerati congiuntamente.

Il risultato è ottimo, in quanto il modello risulta statisticamente significativo sia nei suoi parametri presi singolarmente:

significativo al 99%	0,0003
significativo al 99%	0,00317
significativo al 95%	0,0434
significativo al 99%	0,0002

Sia congiuntamente:

<i>F</i>	<i>Significatività F</i>
11,4962	0,0004

Significatività congiunta del modello testata al 99% di confidenza.

:Per cui, retta di regressione stimata per il nostro modello risulta essere:

$$Y = 80284,98 + 13112,25 * R\&S + 21536,73 * Scambi - 148.388,3 * Costo.Lav.$$

L'incremento di PIL pro-capite, è spiegato nel nostro modello da tre variabili. Esso dipende positivamente dalla presenza delle prime due variabili e negativamente dall'ultima, ossia il costo del Lavoro.

Il coefficiente su cui abbiamo incentrato la nostra analisi, e che risulta essere il più "pesante" in termini assoluti è proprio quello relativo agli scambi col mondo.

A supporto di quest' ultima affermazione consideriamo i seguenti dati:

Pro capite 26002,8768

Costo lavoro 0,472306298

Scambi 0,0230167

Spese in res 1,258126298

Essi sono riferiti all'intero campione italiano, e non alle singole regioni come i dati usati in precedenza per la regressione.

Quindi, nell'ipotesi in cui si incrementasse la percentuale di prodotto interno italiano destinato agli scambi con l'estero, ed aumentasse dell'1%, otterremmo a parità di investimenti destinati a Ricerca&Sviluppo e tenendo fisso il costo del lavoro un impatto sul PIL di:

$$Y = 80284,98 + 13112,25 * R\&S + 21536,73 * Scambi - 148.388,3 * Costo.Lav.$$

$$= 80284,98 + 16495,21 + 22032,29 - 70083,79 = 48728,79 \text{ (1)}$$

Se quindi la media nazionale di prodotto interno percentuale destinato agli scambi esteri dal lato dell'esportazione crescesse, e da 0,0230167 passasse ad 1,02301, a parità di tutto il resto, otterremmo in media, i valori sulla retta (1), e cioè un PIL pro-capite predetto pari a 48728€ circa: la differenza dal PIL di partenza iniziale ed il risultato è quindi di 22725,82€, quasi il doppio!

Ovviamente un incremento di un punto percentuale della variabile “Scambi di merci col mondo” (misurati come %PIL) sarebbe oltre ogni più rosea utopia... Fatto sta che resta impressionante quanto questa variabile incida sul PIL pro-capite nella fattispecie del nostro Paese.

BONTA' DI ADATTAMENTO DEL MODELLO AI DATI

L' R^2 , o coefficiente di determinazione, come illustrato nel capitolo precedente, ci consente di individuare la percentuale di variabilità (devianza) della Y catturata dal modello:

nel caso nostro, ossia quello di un modello multivariato, è necessario considerare anche un'altra variabile che viene riportata sempre sul medesimo output: l' R^2 corretto.

<i>Statistica della regressione</i>	
R multiplo	0,8348
R^2	0,6969
R^2 corretto	0,6363
Errore standard	3838,6059
Osservazioni	19

Come possiamo notare, l' R^2 corretto risulta essere più basso dell' R^2 semplice.

La differenza tra questi due indicatori risiede nel fatto che l' R^2 corretto viene ad essere utilizzato in modelli multivariati.

Più il numero di variabili sale, più troveremo con molta probabilità una differenza sostanziale tra i due indicatori. Tutto tutto per una semplice spiegazione:

L' R^2 corretto è infatti un ulteriore strumento di analisi necessario ad ovviare al problema di eccessivo innalzamento dell' R^2 semplice, non dovuto alla migliore bontà di adattamento, ma al semplice fatto di avere inserito una variabile in più.

L' R^2 corretto infatti penalizza le nuove variabili, mantenendosi su valori più bassi, anche quando ne entrano di nuove a far parte del modello: quello che quindi è il limite principale dell' R^2 semplice, ossia il fatto che cresce per il semplice effetto dell'inclusione di nuove variabili, anche quando esse stesse sono irrilevanti, viene perfettamente ovviato dall' R^2 corretto!

Pur considerando un modello multivariato come il nostro però, la bontà di adattamento ai dati, anche con l' R^2 corretto si mantiene su livelli soddisfacenti:

R al quadrato	0,6363
R al quadrato	0,6969

Volendo interpretare meglio i dati, dato che l' R^2 e R^2 corretto ci dicono se i regressori sono adatti o meno a “spiegare” i valori della variabile dipendente, possiamo con discreta sicurezza affermare che il nostro modello cattura in buona parte la variabilità della Y (PIL pro-capite).

Partendo da un modello di regressione semplice, in cui i ricercatori possono aggiungere o sottrarre altre variabili oltre a quelle del modello di partenza, essi testano via via la significatività dei coefficienti man mano che vengono inclusi.

Nel caso in cui la variabile dipendente nel modello vero dipendesse più variabili, e si andasse a stimare il modello con una sola di esse,

noteremmo di sicuro un R^2 minore di quello che otterremmo dalla regressione di più variabili indipendenti.

Questo perchè quando tralasciamo variabili rilevanti all'interno di un modello, commettiamo un errore di valutazione, i cui effetti si riverberano innanzi tutto sulla bontà di adattamento del modello ai dati: l' R^2 infatti si riduce di molto. Ma anche in termini di coefficienti stimati, in quanto essi risulteranno essere non propriamente pari al valore che si sarebbe ottenuto regredendo per tutte le variabili rilevanti.

La “distorsione” provocata dall'esclusione di variabili rilevanti è detta “Bias” ed ha segno positivo o negativo a seconda che la variabile omessa impatti positivamente o negativamente sul modello.

Nel primo caso illustrato in questo capitolo, abbiamo regredito il Pil Pro-capite” sullo “scambio di merci col mondo”.

Nell'ultimo, quello che abbiamo considerato come “best subset”, abbiamo considerato tre regressori: in termini di R^2 il primo modello spiegava il 12,5% della devianza della Y. Mentre il modello X2 X4 X6 oltre il 60%, e la differenza significativa tra questi due R^2 avvalora quanto precedentemente asserito, e cioè che la bontà di adattamento è funzione del numero di variabili significative incluse nel modello.

OLTRE LA BONTÀ DI ADATTAMENTO DEL MODELLO AI DATI

La bontà di adattamento del modello ai dati, però, non ci informa innanzitutto sulla significatività dei coefficienti considerati, in secondo luogo, non ci dice se il gruppo di regressori è quello migliore di tutti a spiegare i dati reali. In ultima battuta, non ci consente nemmeno di sapere se il modello è o meno specificato correttamente nei suoi parametri.

La presenza o meno di queste tre componenti, imprescindibili in un'analisi di regressione, è stata già precedentemente appurata in ogni passo:

l'analisi effettuata palesava infatti sia la significatività congiunta che singola delle variabili; per quanto riguarda invece il secondo punto, l'individuazione del fascio di variabili che meglio di tutte avrebbero spiegato il modello, veniva affidata al calcolo mediante "best subset", il quale rivelava come il modello che effettivamente è stato considerato fosse il migliore di tutti quelli possibili. Per quel che riguarda la specificazione corretta, trattandosi di dati ottenuti a campione dall'ISTAT per quanto concerne le venti regioni italiane, la consideriamo come pienamente soddisfatta.

Per tutti i motivi che in questo capitolo sono stati illustrati, e per altri che di seguito presenteremo, abbiamo avvalorato e dimostrato, mediante i dati empirici, che il modello di regressione preso in considerazione per la ricerca risulta valido e soddisfacente per i fini preposti. Cerchiamo ora di interpretare i parametri considerati ed arrivare ad alcune conclusioni riguardo quello che potrebbe essere lo scenario economico nazionale futuro nel suo orizzonte di breve periodo.

OSSERVAZIONI E CONCLUSIONI:

L'analisi presentata all'interno di questo lavoro consente possibilità di spunto abbastanza late.

Apartire dalla statistica, passando poi per tutta una serie di materie economiche in generale, per approdare fino alla macroeconomia.

Dei possibili scenari futuri, ci siamo brevemente soffermati all'interno del primo capitolo, ma alla luce delle chiarificazioni scaturite dall'analisi empirica compiuta mediante regressione, possiamo dire di più.

A livello Macroeconomico, l'equazione che comunemente viene ad essere adottata per calcolare per aggregati il reddito nazionale di uno Stato è la seguente:

$$Y = C + G + I + NX$$

Laddove la "Y" rappresenta il prodotto aggregato, la "C" la quota di reddito destinata al consumo, "G" la spesa pubblica, "I" gli investimenti, ed "NX" rappresentano le esportazioni al netto delle importazioni. A giudicare da quest'equazione semplice, ma efficace, è possibile trovare un'interpretazione alla congiuntura negativa che ha coinvolto l'economia nazionale dal 2011, anno a partire dal quale siamo entrati nella recessione.

CONSUMI

La crisi economica, come è stato riportato nell'introduzione a questo lavoro, ha segnato profondamente le abitudini della maggior parte degli italiani in negativo. Le conseguenze della fase di congiuntura negativa, si sono riverberate sul livello dei consumi aggregati. Dall'analisi ISTAT è inoltre emerso che il livello dei consumi rilevato nel 2013 ha fatto un balzo indietro di più di 10 anni, andandosi a posizionare sul livello che era stato raggiunto all'alba del nuovo millennio.

SPESA PUBBLICA:

Le finanze dissestate dello Stato italiano hanno richiesto il tempestivo intervento di un governo tecnico, il quale ha provveduto a risanare le casse dello Stato con una politica fiscale fortemente restrittiva. I livelli di *stock* di debito che erano stati raggiunti negli ultimi anni, hanno inoltre richiesto somme sempre maggiori da destinare alla sterilizzazione degli interessi.

INVESTIMENTI:

Come riportato nel secondo capitolo della relazione annuale 2013, e come trattato sinteticamente nel primo capitolo di questo lavoro, la porzione di prodotto interno italiano destinata agli investimenti, ha visto una fortissima caduta nel settore del credito. La fiducia internazionale minata irrimediabilmente dalla crisi del 2008, ha infatti causato la chiusura dei "rubinetti delle banche", le quali temendo circa l'insolvenza dei loro debitori, si sono trovate costrette ad innalzare spropositatamente le barriere di accesso al credito.

Nella situazione di "*Credit Crunch*" che abbiamo vissuto quindi, si inserisce perfettamente quello che abbiamo sottolineato essere stato il problema che ha condizionato più di tutti la caduta della produttività delle imprese italiane sui mercati esteri, e cioè la mancanza di finanziamenti.

NX:

Le esportazioni nette, che nel nostro modello sono rappresentate dallo “scambio di merci col mondo”, sono l’ultima macroclasse che impatta sul PIL. Nella situazione congiunturale estremamente negativa, in cui il *mixaggio* letale tra crisi economica e crisi del debito sovrano ha portato sull’orlo del baratro la nostra economia, esse hanno contribuito in maniera assolutamente positiva all’aumento del risparmio nazionale, ed in ultima analisi sul PIL stesso.

E GLI SCENARI FUTURI?

Volendo scattare un’istantanea sull’economia italiana oggi, abbiamo una situazione a “luci ed ombre”.

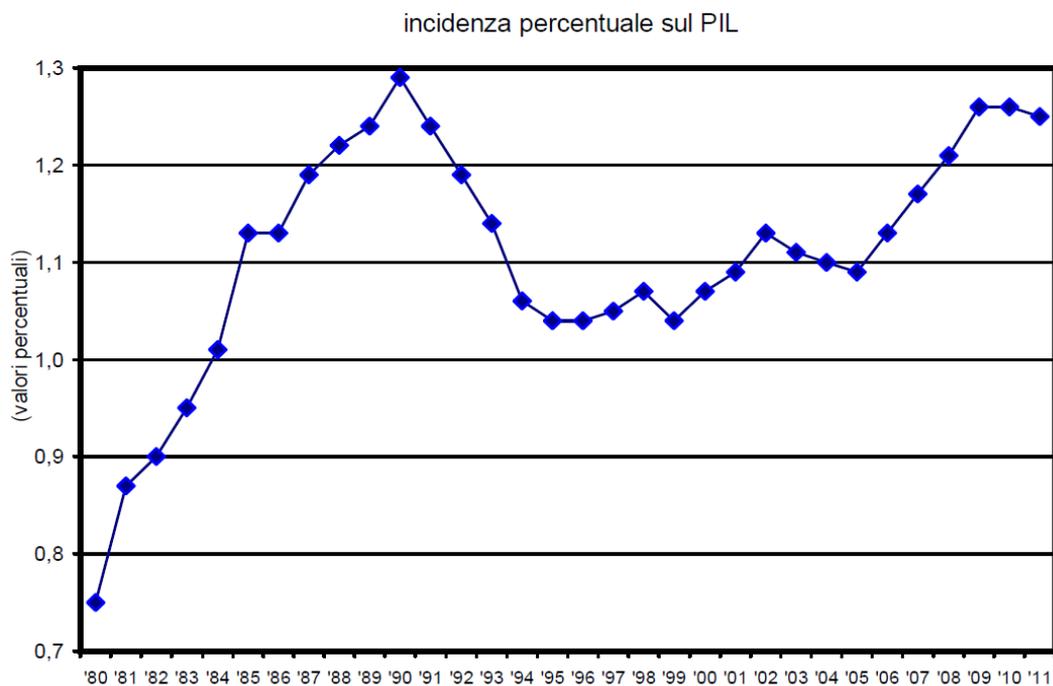
Se nel primo capitolo avevamo definito gli elementi “tenebrosi” che ancora aleggiavano come spettri sulla nostra economia ancora fragile, dall’analisi empirica effettuata possiamo arrivare anche però a delle conclusioni positive.

A livello di finanza pubblica, il peggio sembrerebbe essere passato: il tempo della “*spending review*” e del riordino dei conti a “Tappe Forzate” sta lentamente scemando e sebbene sia ancora troppo presto per potersi lanciare in proiezioni circa le future politiche fiscali che verranno intraprese dai prossimi governi, di sicuro non raggiungeranno mai i livelli di restrizione passati.

Lentamente, quindi sta riprendendo anche la spesa in R&S che dal modello che abbiamo costruito, risulta comunque essere una variabile fortemente rilevante sulla creazione di reddito per gli italiani.

La seguente tabella, ci mostra quella che è la percentuale di reddito nazionale spesa in ricerca e sviluppo:

Fonte AIRI (Associazione italiana per la ricerca industriale)



Nel grafico, il trend negativo inauguratosi a partire da metà del passato decennio, è fortemente correlato alla situazione cosiddetta a “crescita zero”, in cui scarsi investimenti in ricerca e sviluppo, hanno portato a percentuali di crescita del PIL a loro volta scarse.

Secondo le proiezioni però, sembrerebbe che la percentuale di reddito destinata alla spesa in ricerca e sviluppo presente nei budget delle medie e grandi imprese fortunatamente non sia destinata ad essere tagliata.

E LA CRESCITA?

L'economia italiana però è ancora troppo fragile per poter sostenere saggi di crescita a segno positivo. Se nel biennio del governo tecnico la priorità era letteralmente quella di “salvare l'Italia”, a discapito della recessione e della depressione economica, ora che l'Italia è salva, occorrerebbe “salvare gli italiani”, i quali hanno pagato il fio eccessivamente negli ultimi anni in termini di ricchezza ed occupazione...

Occorre quindi spostare il focus sull'economia reale, e sulla capacità delle imprese italiane di riprendere a “macinare” anche all'interno del territorio nazionale. Lo sforzo richiesto al settore industriale è immane: cercare di colmare il *gap* di domanda interna e cercare di riportare i consumi ai livelli pre-crisi.

Per il momento, la ripresa sta partendo dall'estero, ossia dalla capacità “innata” delle nostre imprese di riuscire ad imporsi sui mercati extra italiani con discreta facilità. Tale *skill* delle nostre imprese, dipende ancora fortemente dall'elevata qualità del prodotto, e dal fascino mai tramontato del “*made in italy*”.

Grazie alle esportazioni, le imprese hanno trainato l'economia in recessione alleviando notevolmente le conseguenze della crisi dell'economia reale.

Ma questo non può essere considerato un punto di arrivo:

Tout court, la speranza è quella che in un futuro prossimo, la domanda interna sarà stimolata più di quanto non lo sia oggi, e che i livelli di consumo aggregati si rendano tali da evitare alle imprese italiane di “aggrapparsi” all'estero per vendere i volumi che altresì non raggiungerebbero data la depressione della domanda interna.

Concludendo, sebbene vada contro gli interessi del Sottoscritto, e della ricerca effettuata all'interno di questa tesi di laurea, la speranza per il futuro è quella di non trovarsi più costretti a misurare gli incrementi di PIL pro-capite in funzione delle esportazioni: laddove infatti esse non saranno più la variabile che nel nostro modello era così rilevante, significherà probabilmente che l'economia italiana avrà raggiunto livelli di stabilità sufficienti a considerarci "fuori dai guai".

Significherà probabilmente inoltre, che altri migliori indicatori del benessere diffuso avranno preso il sopravvento: i consumi, il livello di maggiore efficienza dell'istruzione, del *welfare* e via discorrendo...

Allora, potremo dirci soddisfatti, nell'essere riusciti a scongiurare, una volta per tutte, la crisi recessiva in senso lato con la speranza di focalizzare l'attenzione sull'espansione e sulla "*rinascita italiana*", che tutti quanti con ansia attendiamo.

Ringraziamenti:

Non ci saranno mai abbastanza ringraziamenti, e sufficientemente adeguati per tutti coloro, genitori e parenti, che con tanto amore e dedizione hanno profuso, senza spese, nei riguardi del Sottoscritto, chi materialmente, chi “dall’Alto” energie inenarrabili!

Ad essi, ed in particolare alla “sempre presente” fidanzata, sarà riservato sempre un posto speciale nel cuore, per ogni momento intenso, nel bene e nel male, vissuto.

Un ringraziamento sentitissimo, il Sottoscritto desidera che giunga alla Prof.ssa Livia De Giovanni, instancabile ed assidua ricercatrice ed insegnante, la quale si è resa disponibile in ogni istante per far sì che questo lavoro potesse essere portato a buon fine, e che con ardente zelo ha dimostrato la sua eccezionale levatura umana, oltre che quella scientifica.

Al magnifico Rettore, alla Commissione di Laurea, ai Chiarissimi Professori con cui è stato condiviso il triennio giungono

Infinite grazie

Bibliografia:

Anna Clara Monti, "Introduzione alla statistica", Seconda Edizione, Edizioni Scientifiche Italiane.

S.Borra, A. Di Ciaccio - "Statistica, metodologie per le scienze economiche e sociali", Mc Graw Hill.

P. Newbold, W.L. Carlson, B. Thorne, "Statistica", Seconda Edizione, Pearson Prentice Hall.

James H. Stock, Mark W. Watson, "Introduzione all'econometria", Seconda Edizione, Pearson Paravia Bruno Mondadori.

Rapporto Annuale Istat 2013, Capitolo I, II, III:

Devereux M.P. e R. Griffith. 1998. "The taxation of discrete investment choices". Institute for Fiscal Studies, Working Paper, n. 98/16

Istat. 2013b. "Le prospettive per l'economia italiana nel 2013-2014". Roma: Istat. (Previsioni, 6 maggio)

Sace. 2012. "Quando l'export diventa necessario". Rapporto Export 2012-2016, dicembre.