



**UNIVERSITA' DEGLI STUDI DI PISA - DIPARTIMENTO DI
SCIENZE ECONOMICHE**

Tesi di Laurea Specialistica

Asimmetrie informative nel mercato del lavoro: incentivazione
microeconomica con *effort* non verificabile ed effetti
occupazionali

Relatore:

prof. Nicola Meccheri

Candidato:

Gianluca Conte

ANNO ACCADEMICO 2008/2009

“ I salari correnti del lavoro dipendono ovunque dal contratto che comunemente si conclude tra questa due parti i cui interessi non sono affatto gli stessi. Gli operai desiderano ricevere il più possibile, i padroni dare il meno possibile. I primi sono propensi a coalizzarsi per elevare il salario, i secondi per diminuirlo”

Adam Smith (1776)

Indice

Indice	V
Ringraziamenti	IX
Prefazione	XI
1. Analisi del paradigma neoclassico nel mercato del lavoro: ipotesi, sviluppi, limiti e principali implicazioni	1
1.1 Domanda e offerta di lavoro nel modello neoclassico.....	3
1.1.1 La domanda di lavoro.....	3
1.1.2 L'offerta di lavoro.....	9
1.1.3 L'equilibrio di mercato.....	15
1.2 Osservazioni generali e critiche sul modello neoclassico, cenni sull'approccio keynesiano.....	21
1.2.1 L'ipotesi di perfetta simmetria informativa e i limiti della "transazione di mercato".....	27
2. Asimmetrie informative: l'economia dei contratti. L'approccio dei contratti d'agenzia e il problema del rischio	33
2.1 Il <i>trade-off</i> tra incentivo e ripartizione del rischio: dal <i>first</i> al <i>second best</i>	42
2.2 Osservazione di sintesi sui contratti d'agenzia.....	50

Appendici.....	54
A.1 Scelte in condizioni di incertezza: la posizione dei soggetti nei confronti del rischio (cenni).....	54
A.2 L’approccio della teoria dei giochi nella relazione principale– agente: schema semplificato d’analisi.....	59
 3. Salari d’efficienza: le origini della teoria, profilo generale e modello di Shapiro e Stiglitz. Profili salariali legati all’anzianità (cenni)	 69
3.1 Salari d’efficienza: profilo generale, modello di Solow e di Shapiro – Stiglitz.....	73
3.1.1 Schema base nella teoria dei salari d’efficienza: il modello di Solow.....	74
3.1.2 Microfondazione della relazione tra salario ed efficienza: l’ <i>effort</i> come misura dell’efficienza. Modello uniperiodale.....	77
3.1.3 Il modello di Shapiro e Stiglitz.....	80
3.1.4 Considerazioni aggiuntive circa il modello di Shapiro – Stiglitz: spinta della domanda, inclinazione della NSC, efficienza e altre osservazioni.....	88
3.2 Osservazioni di sintesi sul modello di Shapiro e Stiglitz.....	92
3.3 Il sistema delle cauzioni (<i>performance bonding</i>) e i profili salariali legati all’anzianità.....	96
3.3.1 Profili salariali legati all’anzianità: i lavoratori come “creditori” dell’impresa.....	98

4. Incompletezza contrattuale, <i>enforcement</i> endogeno e i contratti <i>self – enforcing</i>: teoria dei contratti impliciti	103
4.1 Contratti <i>self – enforcing</i> : dall'esogeneità all'endogeneità dell' <i>enforcement</i> contrattuale.....	113
4.1.1 <i>Enforcement</i> bilaterale.....	116
4.1.2 <i>Enforcement</i> multilaterale.....	124
4.1.2a Relazione con durata finita e orizzonte infinito per entrambe le parti.....	128
4.1.2b Relazione con durata finita e orizzonte finito per entrambe le parti.....	129
4.2 Osservazioni di sintesi sui contratti <i>self – enforcing</i>	133
5. <i>Rank – order Tournaments</i>: i tornei come schema di incentivazione	137
5.1 Pregi e limiti del torneo: il ruolo delle <i>performance</i> relative, lo <i>spread</i> salariale e il ritorno al <i>first–best</i> . Il modello di Lazear – Rosen.....	141
5.2 Osservazioni di sintesi sui tornei e considerazioni aggiuntive.....	157
6. Analisi degli effetti macroeconomici dei sistemi di incentivazione: la disoccupazione involontaria è un “male” necessario per l'<i>enforcement</i>?	165
6.1 Confronto dei diversi schemi di incentivazione: l'analisi tramite le <i>asset value equation</i>	170
6.1.1 <i>Shirking model</i> , <i>implicit contracts</i> (bonus) e tornei: analisi semplificata sfruttando le <i>asset value equation</i> ..	175
6.1.1a Salari d'efficienza.....	176

6.1.1b	Contratti <i>self-enforcing</i>	177
6.1.1c	<i>Rank-Order Tournaments</i>	183
6.1.2	Analisi degli effetti macroeconomici.....	185
7.	Sintesi della tesi e commenti conclusivi	189
7.1	Sintesi dei diversi schemi di incentivazione.....	191
7.1.1	Contratto d'agenzia: riepilogo.....	191
7.1.2	Salari d'efficienza: riepilogo.....	193
7.1.3	Contratti <i>Self – Enforcing</i> : riepilogo.....	194
7.1.4	<i>Rank-Order Tournaments</i> : riepilogo.....	196
7.1.5	Effetti macroeconomici, conclusioni qualitative: riepilogo.....	197
	Bibliografia	201

Ringraziamenti

Desidero innanzitutto ringraziare il mio relatore, il professor Nicola Meccheri, per i preziosi insegnamenti e per il tempo dedicatomi.

La mia stima per lui è dovuta non solo alla profonda competenza e conoscenza delle materie insegnate, ma anche e soprattutto alla profonda umanità con cui si relaziona ai propri studenti e con cui mi ha saputo tranquillizzare nei momenti di maggiore incertezza.

Uno speciale ringraziamento va al prof. Fabrizio Bientinesi che in più di un'occasione, nel corso della mia carriera di studente universitario, ha saputo incoraggiarmi e sostenere.

Doveroso e sentito è anche un ringraziamento a tutti i professori che ho avuto, in particolar modo quelli del dipartimento di Scienze Economiche, i quali, oltre a farmi crescere intellettualmente, hanno contribuito in modo significativo alla mia crescita umana. In particolare prof. Enrico Ghiani, prof. Paolo Scapparone, prof. Luca Spataro, prof. Carlo Bianchi.

Un affettuoso ringraziamento anche a tutti i colleghi conosciuti nel corso di questi anni, con i quali ho condiviso paure, aspirazioni e momenti indimenticabili. In particolare Daniele, Alessandra, Filippo, Mary, Cristina, Giulia, Marco, Francesca, Simona “sarda”, Simona “catanese”, Omar, Gianni, Chiara, Luca, Andrea.

Desidero ringraziare anche i miei carissimi amici Valerio, Tommaso, Matteo, Filippo e Manuela, i quali hanno sempre creduto in me.

Infine, ma non certo per ordine di importanza, voglio ringraziare la mia famiglia; mia sorella maggiore che è sempre riuscita a “sopportarmi” ed incoraggiarmi in modo incessabile e i miei genitori a cui dedico questo lavoro che, in fondo, è tanto mio quanto loro.

Prefazione

L'economia moderna è divenuta, grazie in particolare alla rivoluzione marginalistica, una scienza a tutti gli effetti capace di “competere” sul piano metodologico anche con le scienze considerate “perfette”.

L'aspetto metodologico in economia, pur giocando un ruolo assolutamente centrale, raramente trova un adeguato spazio d'analisi: lo studio dei vari paradigmi economici, ovviamente, è imprescindibile da considerazioni di carattere metodologico, tuttavia, con un po' di sorpresa, ho notato, nel corso della mia carriera da studente universitario, che agli aspetti relativi al metodo scientifico in economia non viene attribuito un adeguato spazio.

Quelle che seguono sono delle semplicissime considerazioni che ho voluto inserire alla luce del mio personale interesse relativo al “metodo-economico/scientifico”, considerazioni lontane dall'essere esaustive.

Lo strumento idoneo all'interpretazione della realtà è un sistema formale, un modello. Attraverso la modellazione si ha una semplificazione della realtà e di conseguenza sia ha anche una semplificazione dell'analisi; ci vediamo costretti, però, a rinunciare a pretese di assolutismo razionalistico, poiché, come mostrato da Kurt Gödel, ogni sistema formale o informale di rappresentazione è per costruzione incompleto, o meglio: qualsiasi sistema formale, compreso anche il linguaggio, ad esempio, oscilla tra coerenza e completezza, non permettendo di giungere contestualmente ad entrambi gli obiettivi.

Risulta evidente, quindi, che la “ricerca”, anche in ambito economico, sarà necessariamente “limitata” in quanto, come già accennato, non viene effettuata sulla realtà, ma sul modello di tale realtà.

Nel corso della storia dell'economia si sono succeduti economisti convinti dell'esclusiva utilità dell'analisi applicata ed economisti, invece, convinti che l'analisi teorica, fosse ugualmente, se non maggiormente, importante.

Le scienze economiche sono nate nel settecento ed il primo approccio utilizzato è stato quello comune a tutte le altre scienze, ossia quello naturalistico: inizialmente l'oggetto d'analisi era il semplice scambio e la produzione.

Tale era anche l'approccio del padre dell'economia politica, Adam Smith: in base all'impostazione dell'economista scozzese, l'economia era puro egoismo e veniva pertanto guardata con timore e sospetto ma accettata come male necessario; la sfera morale sembrava assumere un'importante ruolo ma in realtà era lo stesso Smith a sottolineare che la dicotomia "interesse individuale (egoismo) e interesse pubblico", di fatto, non esiste e che è l'egoismo di ciascuno di noi a far andare avanti la società.

Ben presto, però, l'approccio metodologico nelle scienze, e quindi anche in economia, subì notevoli cambiamenti: dapprima non ci si domandava cosa fosse la scienza, tuttavia rispondere a questa domanda era divenuto ormai necessario per poter procedere in qualsiasi campo di ricerca.

Dal positivismo di Comte e Mill si è passati alla "scienza astratta funzionale" di Poincaré, per giungere poi al neopositivismo, criticato successivamente sia sul piano fisico (principio di indeterminazione di Heisenberg) sia sul piano sociale (Popper) sia sul piano matematico/formale (Godel).

La scuola di sintesi neoclassica nasce proprio dietro l'impulso dell'impostazione neopositivistica, nel tentativo di avvicinare lo studio delle relazioni economiche a quello delle scienze pure. Il nuovo paradigma non fu esente da critiche, in particolare molti studiosi ritenevano eccessiva la semplificazione dei soggetti economici e rivendicavano una maggiore complessità della dimensione umana.

All'interno del paradigma neoclassico è opportuno, ad esempio, ricordare l'enorme contributo, anche metodologico, di Lèon Walras nell'ambito della concorrenza perfetta: l'economista francese era convinto del fatto che, pur nell'ambito di una completa autonomia metodologica, l'analisi teorica fosse chiamata a svolgere un ruolo di guida nelle scelte relative alla concreta organizzazione dei mercati¹.

Il passo metodologico successivo al paradigma neoclassico prevedeva un approccio non più basato sulle relazioni tra i soggetti ma piuttosto sull'ottimizzazione individuale di risorse scarse per usi alternativi.

Il termine "ottimizzare", però, implica l'individuazione certa e generale di tutte quelle variabili che accrescono o riducono l'utilità dei soggetti, si parla di "razionalità", ma è razionale etichettare in modo "generalista" ciò che, economicamente, è un "bene" e cioè che è un "male"?

Sembrerebbe, quindi, di essere giunti ad un momento in cui l'economia possa essere studiata solo tramite una rappresentazione astratta della realtà, tramite un modello: nonostante le divergenze d'opinione in merito ad ipotesi più o meno realistiche, tutte le moderne scuole di pensiero, a partire anche dai neoclassici, utilizzano un sistema formale.

Nell'elaborare un modello economico non dobbiamo, pertanto, pretendere che il modello stesso spieghi "tutto" (compreso ciò che è al di fuori della nostra analisi) cioè sia "perfetto", in quanto questo non è, e non deve essere, l'obiettivo; il modello deve essere sviluppato su ipotesi a priori (suggerite dai dati ma anche da intuizioni personali), e non vi è la stretta necessità che queste ipotesi siano completamente realistiche.

¹ Pavanelli G., *Valore, distribuzione, moneta. Un profilo di storia del pensiero economico*, FrancoAngeli. 2003, pg. 196

Tali ipotesi, infatti, possono essere modificate successivamente a seguito, ad esempio, di una generalizzazione del modello. E' possibile, inoltre, che il modello, pur partendo da ipotesi non realistiche conduca a risultati "applicabili" anche nella realtà: ad esempio l'analisi della concorrenza perfetta di Walras, pur non avendo applicabilità reale diretta, ha suggerito determinati comportamenti per migliorare i risultati.

Negli ultimi anni c'è stato un impiego sempre maggiore e pervasivo di strumenti matematici all'interno dell'analisi economica, questo ha consentito di sviluppare nuovi settori di ricerca come ad esempio la teoria dei giochi o l'econometria e ha consentito inoltre di dare rigore formale a intuizioni e idee che altrimenti non avrebbero avuto seguito.

In tal contesto, l'economia dei contratti è un significativo esempio: la relazione tra le parti, il rischio di opportunismo, i tentativi di indurre le parti ad essere oneste sono sempre stati oggetto di speculazioni in vari ambiti, tuttavia non è stato mai possibile dare rigore formale a tali relazioni; questo limite ha impedito un'analisi coerente, precisa ma soprattutto generale del problema.

Senza dare rigore formale a concetti come preferenze, utilità, reputazione, la teoria dei contratti non sarebbe mai esistita e avremmo avuto solo "speculazioni" informali.

Anche nell'analisi economica è, quindi, di fondamentale importanza l'elaborazione di una teoria, di un modello, che rappresenti non il punto di arrivo dell'analisi, ma quello di partenza.

Per concludere, sembra adeguato citare l'astrofisico Stephen Hawking, il quale fornisce una definizione "universale" del concetto di Teoria, applicabile non solo in fisica, ma in tutte le scienze conoscitive:

"una teoria è una buona teoria se soddisfa due condizioni: deve descrivere accuratamente un'estesa serie di osservazioni sulla base di un modello che contiene solo pochi elementi arbitrari, e deve fare predizioni precise riguardo ai risultati di osservazioni future" ².

Dopo questo brevissimo *excursus* metodologico, passiamo a introdurre gli obiettivi e le finalità della tesi. L'analisi che ho portato avanti in questo lavoro ha due principali punti di riferimento: la problematica del *moral hazard* nell'ambito del mercato del lavoro e gli effetti macroeconomici dei diversi schemi di incentivazione, volti a eliminare tale problema.

L'analisi che intendo svolgere ha, quindi, sia una dimensione che attiene strettamente alla sfera microeconomica, sia una dimensione che, invece, riguarda essenzialmente aspetti macroeconomici.

Nel modello neoclassico del mercato del lavoro, l'ipotesi di informazione perfetta e simmetrica (congiunta in particolare³ a quella di perfetta flessibilità salariale) non lascia spazio al fenomeno della disoccupazione involontaria.

La letteratura successiva, però, si è concentrata sull'ipotesi più realistica che l'informazione non fosse distribuita in modo perfetto e simmetrico tra tutti i soggetti e che le rigidità salariali registrate nel mercato potessero essere microfondate (e quindi non derivassero solo da imperfezioni del mercato).

Abbandonato il paradigma neoclassico, quindi, è stato possibile creare un filone di ricerca volto a spiegare le rigidità salariali e la presenza (nonché la persistenza) di disoccupazione involontaria.

² Hawking W.S., *Breve storia del tempo – Dal Big bang ai buchi neri*, Rizzoli, p 23. 1989

³ I risultati del modello neoclassico si basano anche su altre importanti ipotesi, oggetto d'analisi nel capitolo 1.

In altre parole, l'analisi delle asimmetrie informative, pur partendo da basi microeconomiche, in un primo momento fu orientata quasi interamente allo studio di aspetti macroeconomici legati al mercato del lavoro.

Per quanto riguarda, invece, gli aspetti prettamente microeconomici, le asimmetrie informative comportano il rischio d'opportunismo contrattuale: un soggetto ha maggiori informazioni rispetto al partner e cerca di trarne il massimo vantaggio (anche a scapito di altri soggetti coinvolti nella transazione).

In questo lavoro viene identificato e delineato il problema di fondo, cioè il problema della relazione tra due parti aventi interessi contrapposti (in un contesto lavorativo), e vengono analizzate, a livello microeconomico, i principali sistemi di *enforcing* contrattuale (sistemi, cioè, che garantiscono l'effettivo rispetto degli accordi) volti a risolvere tale contrapposizione.

Il punto fondamentale da chiarire è il seguente: in origine l'analisi microeconomica delle asimmetrie informative (in particolare i salari d'efficienza e la "prima generazione" dei contratti impliciti) era interamente volta e funzionale agli aspetti macroeconomici e, in questo filone, riusciva a fornire una buona spiegazione della disoccupazione involontaria tramite la microfondazione del così detto "pavimento salariale".

Con lo sviluppo della teoria dei contratti, però, sono stati approfonditi anche aspetti puramente microeconomici relativi all'incentivazione lavorativa, senza "addentrarsi" (almeno inizialmente) nell'analisi degli effetti aggregati.

Questi nuovi filoni (in particolare i contratti impliciti della "seconda generazione"⁴ e i tornei) hanno consentito di ottenere risultati "equivalenti" (in termini di incentivi lavorativi) dei modelli che microfondano la rigidità salariale

⁴ Intesi, cioè, nel senso di contratti *self-enforcing*.

senza, però, implicare necessariamente essi stessi disoccupazione involontaria (o quantomeno implicandone un livello inferiore).

L'obiettivo finale di questo lavoro consiste nell'analizzare gli effetti macroeconomici di questi schemi, analisi che si soffermerà in particolare sul confronto degli effetti a livello aggregato dei tre più significativi filoni di *enforcing*, ovvero i salari d'efficienza, i contratti *self-enforcing* e i tornei. In sostanza, si risponderà alla domanda se è necessaria o meno la disoccupazione per annullare il rischio di opportunismo individuale.

La risposta, come vedremo, sarà negativa⁵; infatti, è possibile individuare schemi "microeconomici" di incentivazione senza produrre disoccupazione a livello aggregato.

Ovviamente la realtà è estremamente sfaccettata e non può essere "rinchiusa" in uno o in un altro modello; tuttavia, alla luce di quanto detto, sembrerebbe ragionevole dover approfondire il problema della disoccupazione in altre direzioni, come ad esempio, verso la teoria della concorrenza imperfetta, basata sull'idea che le parti abbiano un forte potere di mercato.

Nella prima parte della tesi viene introdotto il modello neoclassico del mercato del lavoro, caratterizzato da simmetria informativa, flessibilità salariale e concorrenza perfetta; la presentazione di questo mercato è necessaria per comprendere i successivi sviluppi che hanno portato alla microfondazione delle rigidità salariali. In un certo senso, quindi, il mercato neoclassico funge da modello *benchmark*.

Nella seconda parte della tesi, viene introdotta la presenza di asimmetrie informative tra impresa e lavoratore e viene descritta la problematica principale

⁵ Questo dipende, ovviamente, dalle assunzioni adottate: in generale l'aspetto rilevante è il grado d'efficienza sociale relativa ottenibile da ciascuno schema.

– agente nell’ottica del contratto d’agenzia; successivamente vengono descritti, a livello microeconomico gli schemi di incentivazione basati sui salari d’efficienza, sui contratti *self-enforcing* e sui tornei.

L’ultima parte della tesi è dedicata al confronto degli effetti macroeconomici dei vari sistemi di *enforcing*. Viene mostrato che la disoccupazione, prevista dai salari d’efficienza, non è sempre necessaria per ottenere l’*enforcing* contrattuale: i contratti impliciti, in particolare, pur generando disoccupazione (a causa di problemi legati alla credibilità dell’impresa) ne generano un livello inferiore rispetto ai salari d’efficienza, inoltre sotto date condizioni la disoccupazione è volontaria; i tornei, invece, consentono di incentivare i lavoratori senza la necessità di creare disoccupazione⁶.

⁶ Questo avviene, in particolare, supponendo che l’impresa possa valutare con precisione l’impegno individuale. Abbandonando questa ipotesi, come vedremo, i tornei restano il sistema di incentivazione maggiormente efficiente ma implicano essi stessi disoccupazione involontaria.

1. Analisi del paradigma neoclassico nel mercato del lavoro: ipotesi, sviluppi, limiti e principali implicazioni

Ai fini dell'analisi che intendo svolgere all'interno di questo lavoro, è opportuno, preliminarmente, chiarire e definire alcuni aspetti fondamentali delle teorie economiche antecedenti agli approcci che hanno portato alla moderna letteratura scientifica relativa alla microfondazione delle rigidità salariali.

Per poter comprendere a pieno le molte¹ teorie riguardanti la rigidità salariale, volte a spiegare fenomeni di disoccupazione prolungata, risulta opportuno, quindi, andare a vedere in primo luogo quali sono state le principali linee guida del paradigma neoclassico², all'interno del quale, come vedremo, non vi può essere disoccupazione involontaria, in particolare grazie all'ipotesi di perfetta flessibilità di prezzi e salari. Il mercato del lavoro neoclassico viene utilizzato, in altre parole, come termine di partenza e paragone per meglio comprendere gli sviluppi successivi.

Gli economisti neoclassici vedevano il mercato del lavoro come qualsiasi altro mercato. L'ipotesi di base era rappresentata dal contesto di perfetta concorrenzialità che veniva introdotta come presupposto fondamentale.

L'impostazione neoclassica prevede una domanda di lavoro negativamente inclinata, in quanto espressione della produttività marginale del lavoro, e un'offerta positivamente inclinata, espressione della disutilità marginale del lavoro. L'intersezione tra domanda e offerta dà l'equilibrio (E) del mercato caratterizzato dall'assenza di disoccupazione involontaria: la disoccupazione,

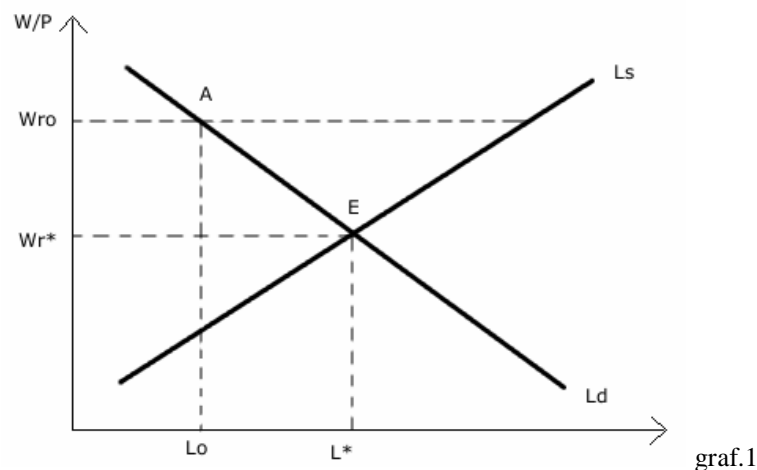
¹ Molti sono stati i contributi teorici finalizzati a microfondare le rigidità salariali; oltre a quelli che verranno analizzati in questo lavoro vanno ricordati, ad esempio, i modelli insider-outsider, di disequilibrio e di concorrenza imperfetta

² La scuola neoclassica nasce nell'ultimo trentennio dell'800 e si identifica, almeno nel suo sviluppo iniziale, con la scuola marginalista

oltre quella volontaria, che però si registrava empiricamente, veniva classificata come frizionale, derivante cioè dalle imperfezioni del mercato.

Ad un livello di salario reale più alto di quello d'equilibrio (W_{ro}), ci sarà un eccesso di offerta che genera disoccupazione involontaria.

Per gli economisti di orientamento neoclassico, data la situazione di eccesso d'offerta, la pressione competitiva tra i lavoratori, porterà il sistema nuovamente all'equilibrio dove valgono sia il primo che il secondo postulato neoclassico, dove cioè il salario reale è uguale alla produttività marginale del lavoro (primo postulato) ed è uguale anche alla disutilità marginale del lavoro (secondo postulato)



La dinamica di aggiustamento così descritta necessita, per funzionare, della perfetta flessibilità salariale, grazie alla quale è sempre possibile giungere nuovamente alla situazione di equilibrio, nonché della perfetta simmetria informativa.

1.1 Domanda e offerta di lavoro nel modello neoclassico

Domanda e offerta di lavoro dipendono dal salario reale: la domanda, negativamente inclinata indica che le imprese assumeranno un numero maggiore di soggetti quanto minore è il salario reale; l'offerta di lavoro, positivamente inclinata, indica che al crescere del salario reale ci sarà un numero maggiore di soggetti disposti a lavorare. Entriamo a questo punto nel dettaglio della costruzione della domanda e dell'offerta.

1.1.1 La domanda di lavoro

Consideriamo in primo luogo la domanda di lavoro; l'impostazione neoclassica è di breve periodo, quindi, ipotizzata l'esistenza di un'impresa che produce utilizzando capitale e lavoro, sarà possibile modificare il numero dei lavoratori ma non il livello del capitale che consideriamo, pertanto, costante. Questo è un aspetto fondamentale dell'impostazione neoclassica, infatti, come vedremo, la produttività dipende solo dal contesto tecnologico e questo contribuirà ad implicare l'assenza di disoccupazione involontaria.

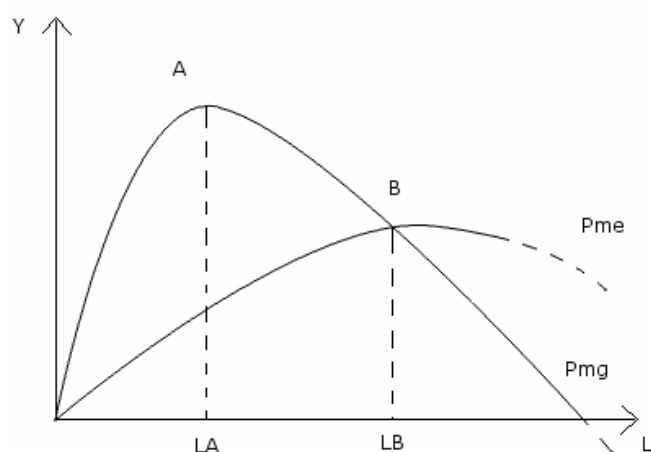
Indichiamo con $Y = F(K, L)$ la funzione di produzione³ dell'impresa, ossia una relazione tecnologica che in aggregato può essere vista come la relazione tra occupazione e prodotto: L è il lavoro impiegato, ipotizzato omogeneo, e K è lo stock di capitale che nel breve periodo è fisso. Ipotizzare che il capitale sia fisso nel breve periodo è un'ipotesi del tutto plausibile alla luce dei tempi e delle modalità di ammodernamento dell'impresa, quindi, la tecnologia, nel breve periodo, è anch'essa data e immutabile.

Ipotizziamo che una quantità crescente di lavoro impiegato si traduca in una quantità crescente di prodotto; questa relazione è la produttività del lavoro che

³ Brucchi L., *Manuale di economia del lavoro*, Il mulino, p 78. 2001

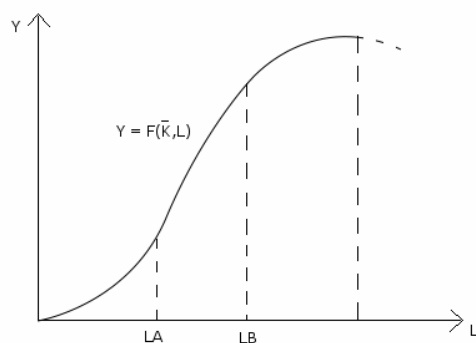
può essere distinta in produttività marginale e produttività media: la prima indica quale è il contributo, in termini di output, di un'unità aggiuntiva di lavoro, cioè il contributo del lavoratore marginale, mentre la seconda è data dal rapporto tra output e il numero dei lavoratori (Y/L).

Per quanto riguarda l'andamento della funzione di produzione, viene comunemente ipotizzato che per bassi livelli di occupazione l'impiego di un'unità aggiuntiva di lavoro produca un elevato livello di output, mentre per livelli di occupazione elevati, introdurre un ulteriore lavoratore produce una variazione dell'output positiva, ma "bassa". Date queste ipotesi si può rappresentare graficamente⁴ l'andamento della produttività marginale e media, entrambe le quali hanno un tratto crescente ed un tratto decrescente



graf.2

Data questa rappresentazione della produttività dei lavoratori, in corrispondenza del punto A, la funzione di produzione presenterà un flesso: la funzione di produzione sarà cioè convessa fino al livello d'occupazione L_A , per poi diventare concava. Graficamente avremo

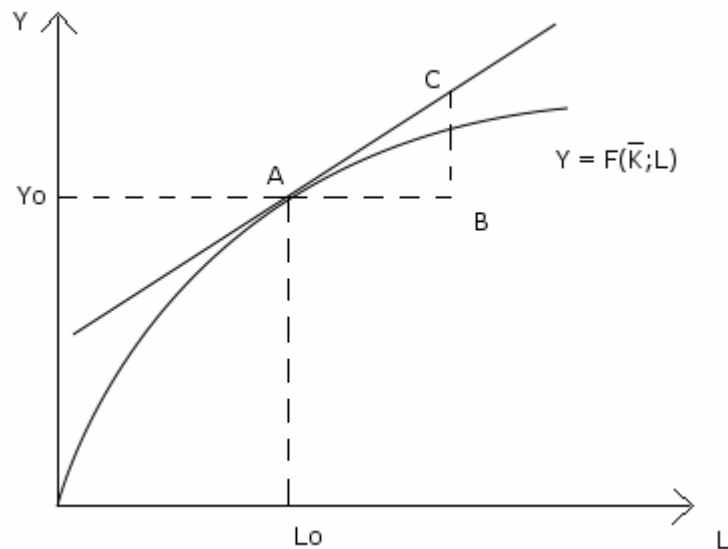


graf.3

⁴ Brucchi L., *op.cit.*, p 79

Per semplificare l'analisi è possibile considerare solo il tratto concavo della funzione di produzione, infatti è plausibile ipotizzare che sul tratto convesso ci si muoverà solo durante i primi periodi in cui l'attività produttiva viene avviata.

Rappresentiamo solo il tratto crescente della la funzione di produzione:



graf.4

Come precedentemente affermato i rendimenti sono decrescenti, cioè la produzione cresce meno che proporzionalmente rispetto ad un aumento dell'occupazione: aggiungendo nuovi lavoratori, l'apporto di questi ultimi, in termini di aumento della produzione, è sempre positivo ma via via sempre più basso in quanto il capitale è fisso e la produttività dipende solo dalle abilità innate dei lavoratori e dalla tecnologia d'impresa.

Dato lo stock di capitale, infatti, è dato anche il livello tecnologico: supponiamo che all'interno di un'impresa ci sia un certo numero e una certa tipologia di macchinari e di strumenti tecnologici finalizzati alla creazione dell'output; avviando l'attività produttiva, l'apporto di ciascun nuovo lavoratore sarà "fondamentale" in quanto andrà a soddisfare delle esigenze "contingenti"; al crescere del numero di soggetti impiegati, la "necessità" di lavoratori per il funzionamento della tecnologia è via via più bassa fino ad arrivare "virtualmente" ad una situazione in cui, dato un macchinario che necessita di x persone per funzionare, il lavoratore aggiuntivo "intralcia" il lavoro altrui.

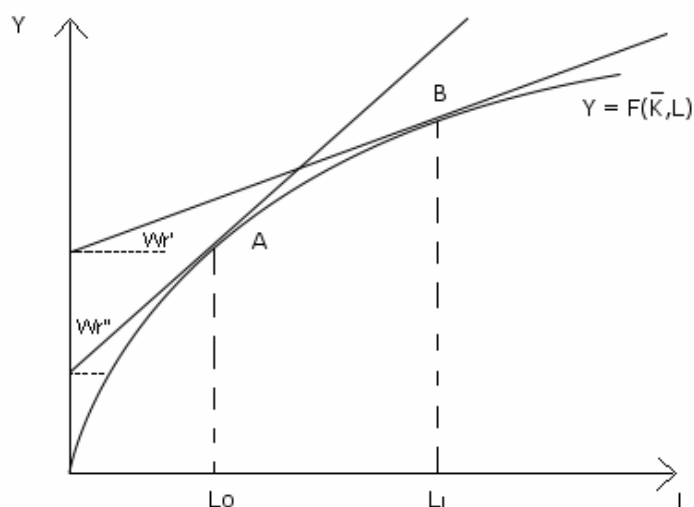
L'ipotesi neoclassica di rendimenti decrescenti, quindi, deriva strettamente dalla dimensione temporale di breve periodo; nel lungo periodo, infatti, è possibile modificare la tecnologia influenzando anche sulla produttività dei lavoratori.

Il contributo marginale di un aumento dell'occupazione è indicato dal rapporto tra il segmento BC e il segmento AB; contributo marginale che è chiamato anche produttività marginale del lavoro.

Le imprese, in quanto soggetti razionali, hanno come obiettivo la massimizzazione dei profitti che corrispondono alla differenza tra ricavi totali e costi totale: dato P il prezzo di vendita unitario e W il costo unitario, cioè il salario unitario dato che nel breve non ci sono costi legati al capitale, i profitti sono dati da

$$\Pi = PY - WL \quad (1.1)$$

e il livello di occupazione è ottimo solo quando si ha l'eguaglianza tra la produttività marginale del lavoro in termini di valore, che rappresenta il ricavo "marginale", e il costo aggiuntivo "marginale" pari al salario unitario⁵; quindi al livello ottimale di L si avrà $F'(K,L) = W/P = W_r$ (salario reale), condizione che graficamente corrisponde, ad esempio, ai punti A e B nel grafico 5



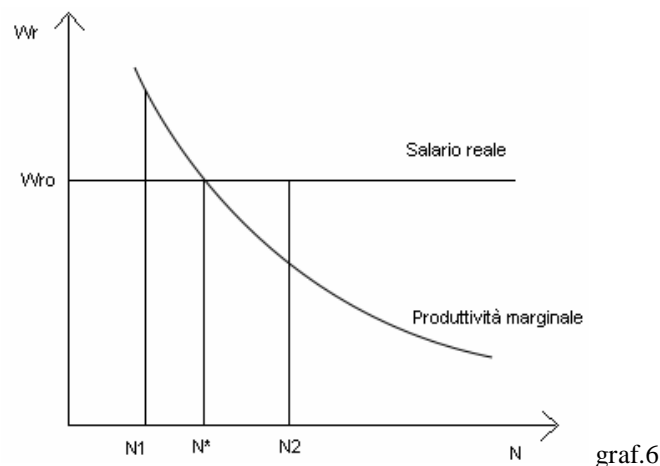
graf.5

⁵ In termini analitici $F'(K,L)P = W$, da cui $F'(K,L) = W/p = W_r$.

Da questa rappresentazione vediamo che la condizione di ottimo è soddisfatta per due diversi livelli di salario reale; il livello associato ad una minore occupazione (L_0) è caratterizzato da un salario reale più grande (W_r'') rispetto al livello associato ad una maggiore occupazione (L_1): i livelli di occupazione L_0 e L_1 (graf.5) appartengono alla funzione di domanda di lavoro

A partire dalla funzione di produzione è possibile ricavare la domanda stessa, le imprese, infatti, continueranno ad assumere lavoratori fino al momento in cui il contributo del lavoratore marginale in termini di valore è pari al costo marginale del lavoro, cioè il salario: come visto in precedenza, un'impresa assumerà un lavoratore fin quando il contributo di quest'ultimo in termini di produzione è maggiore rispetto ai costi che l'impresa deve sostenere pagando il salario.

Il contributo del lavoratore è indicato dalla produttività marginale che è la pendenza della funzione di produzione: assumendo un lavoratore, l'impresa godrà di un aumento del prodotto pari alla produttività marginale e dovrà sostenere un costo pari al salario reale; ne consegue che le imprese assumono lavoratori fino all'ottenimento dell'uguaglianza tra beneficio e costo marginale, graficamente

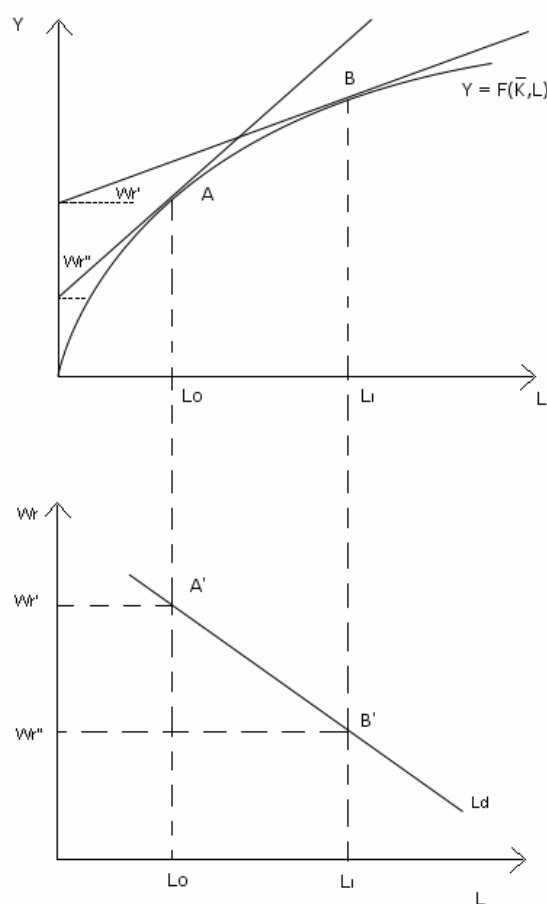


Se l'impresa impiega N_2 (graf.6) graficamente vediamo che il salario reale è maggiore rispetto la produttività: l'impresa sta impiegando troppo lavoro; abbassando l'occupazione si avrebbe una riduzione anche della produzione e

quindi dei ricavi, ma avremo anche una riduzione dei costi maggiore rispetto alla riduzione dei ricavi. Una dinamica speculare si ha al livello N1.

L'unico livello di occupazione che consente un bilanciamento è N^* , dato il livello del salario reale: la domanda di lavoro da parte di una impresa coincide, quindi, con la sua curva della produttività marginale.

Dal confronto tra i punti A e B del grafico 7 risulta evidente che la curva di domanda di lavoro è negativamente inclinata e decresce al crescere del salario reale, graficamente



graf.7

da questo grafico è possibile desumere un aspetto molto importante dell'impostazione neoclassica, ossia che la curva di domanda di lavoro dipende dalla concavità della funzione di produzione, questa considerazione è fondamentale in quanto consente di capire che la domanda di lavoro neoclassica incorpora la tecnologia dell'impresa.⁶

⁶ Ibid, p 80

1.1.2 L'offerta di lavoro

Per quanto riguarda, invece, la curva d'offerta di lavoro, come visto in precedenza, se il salario reale aumenta, tra la forza lavoro ci sarà un numero maggiore di soggetti disposti ad offrire il proprio lavoro per un dato livello di salario e quindi l'offerta sarà positivamente inclinata.

L'idea alla base della curva d'offerta di lavoro neoclassica può esser fatta risalire a Stanley Jevons, uno dei “padri fondatori” del marginalismo, anche se in realtà la formulazione di una relazione che indica la quantità di lavoro erogata sul mercato in funzione del salario verrà sviluppata successivamente.

Jevons si “limita” a introdurre una teoria marginalistica della scelta, ma ormai era già chiaro quale sarebbe stato il *leit motiv* della produzione scientifica successiva.

I lavoratori hanno una propria funzione di utilità⁷ in base alla quale prenderanno le proprie decisioni; data l'ipotesi di razionalità, ciascun soggetto massimizzerà la propria funzione di utilità e il problema di massimizzazione può essere impostato in questi termini⁸

$$\max_{(C,T)} U(C,T) \quad (1.2)$$

$$\text{s.to} \quad PC + WT \leq WT_{\max} + X \quad (1.3)$$

Gli argomenti della funzione di utilità sono C e T, rispettivamente il consumo (realizzabile grazie all'ottenimento del reddito da lavoro) ed il tempo libero⁹, entrambi contribuiscono nello stesso verso all'utilità dei soggetti e pertanto all'aumentare di uno è necessario che l'altra componente venga ridotta per avere lo stesso livello di utilità. La massimizzazione è vincolata e W, T_{max} e X sono

⁷ Ipotezzata continua, derivabile e strettamente quasi concava.

⁸ *Ibid.*, p 62

⁹ Il salario nominale è il costo opportunità del tempo libero, è ciò che il lavoratore non ottiene se decide di non lavorare, in sostanza il salario è il prezzo del tempo libero

rispettivamente il salario orario, il tempo totale a disposizione di ciascun soggetto e i redditi non da lavoro; il tempo che il lavoratore dedica al lavoro è dato da $L = T_{\max} - T$ ed il vincolo, rappresentato nel piano di coordinate C e T (graf.9) è la retta con inclinazione negativa $-W/P$ e interseca l'asse verticale al livello del reddito non da lavoro. Considerando la (1.3) come un'eguaglianza è possibile, infatti, ricavare

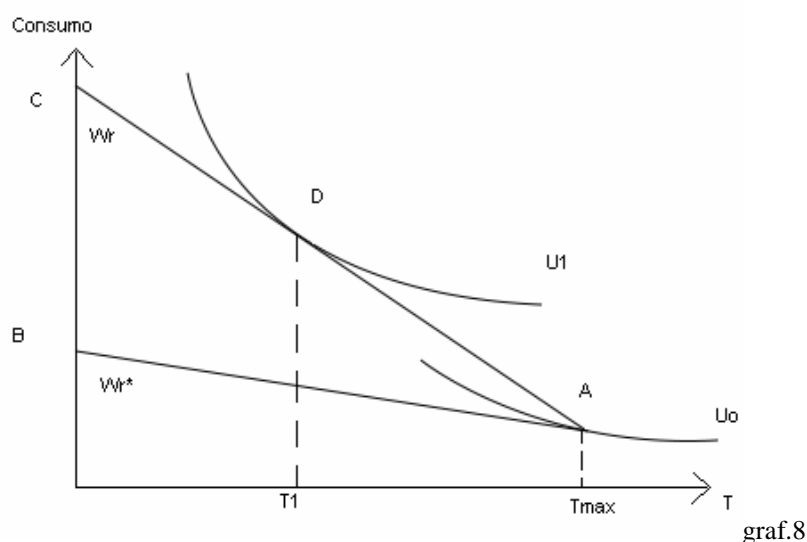
$$PC = W(T_{\max} - T) + X \quad \text{da cui} \quad (1.4)$$

$$C = (WT_{\max} + X)P^{-1} - (W/P)T \quad (1.5)$$

da cui si vede che l'inclinazione del vincolo è data da $-(W/P)$ che per semplicità possiamo indicare con W_r , cioè il salario in termini reali.

All'interno del vincolo, l'espressione $(WT_{\max} + X)P^{-1}$ è detta reddito potenziale “pieno” che si ottiene qualora il lavoratore dedicasse tutto il suo tempo al lavoro.

Per capire meglio alcuni fondamentali aspetti della scelta dei soggetti è opportuno analizzare il problema da un punto di vista grafico¹⁰,



La retta AB (graf.8) rappresenta il vincolo di bilancio più “basso” e con la pendenza minima (cioè $-W_r^*$) ed A rappresenta una soluzione ad angolo: nel

¹⁰ *Ibid*, p 63. Ipotizzando, in particolare, che il salario alternativo sia positivo.

punto A si vede quale è il livello di consumo di un soggetto che non lavora (infatti T_{max} è interamente dedicato al tempo libero); l'inclinazione della retta AB è data dal così detto salario di riserva al quale il soggetto è del tutto indifferente se lavorare o meno.

Consideriamo ora un altro livello di salario, il livello W_r (graf.8); data l'ipotesi di razionalità, il soggetto lavorerà nel momento in cui il salario è strettamente maggiore rispetto a quello di riserva che può essere visto come il rapporto tra l'utilità marginale del tempo libero e l'utilità marginale del consumo, in termini analitici

$$W_r = \frac{W}{P} = \frac{\frac{\partial U}{\partial T}}{\frac{\partial U}{\partial C}} \quad (1.6)$$

Dalla massimizzazione vincolata, l'individuo sceglierà il punto D (graf.8) (tangente tra le curve di indifferenza e il vincolo) che definisce la combinazione ottima tra tempo libero e consumo che massimizza l'utilità del soggetto.

In sostanza, dal grafico precedente, a partire dalle preferenze individuali di ciascun lavoratore, è stato determinato un livello di salario reale sotto il quale il soggetto non offrirà il proprio lavoro.

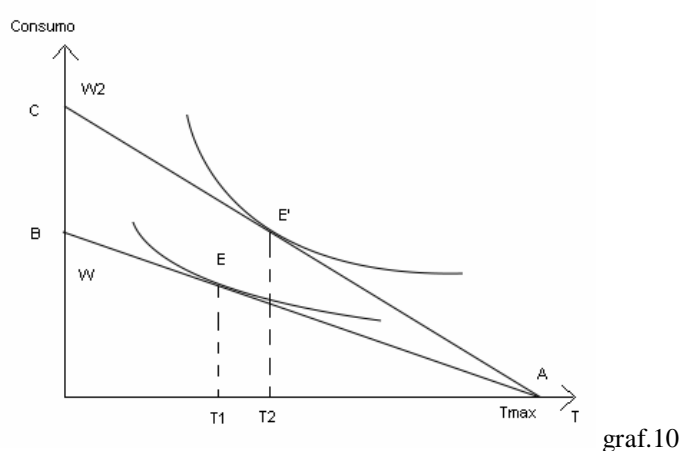
In questa situazione, a parità di prezzi, un aumento salariale genera due effetti contrapposti:

- si ha un aumento del costo opportunità del tempo libero e questo spinge i lavoratori ad offrire più lavoro di più (effetto sostituzione)
- a parità di quantità di lavoro, l'aumento del salario monetario determina un aumento del reddito e, quindi, se il tempo libero è un bene normale avremo una riduzione dell'offerta individuale di lavoro (effetto reddito)

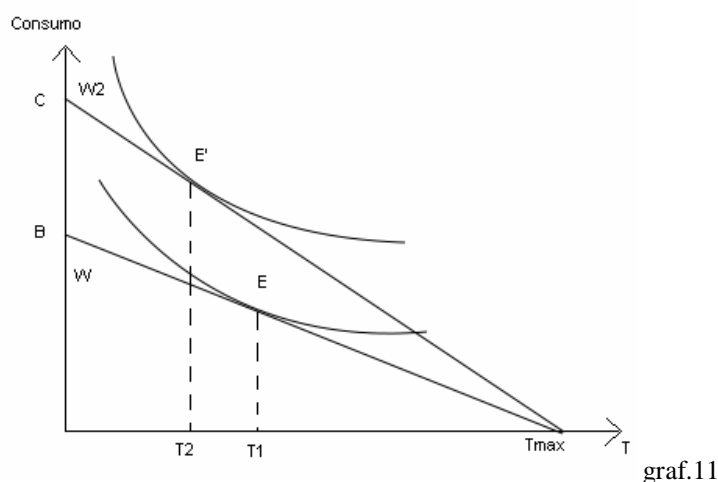
La prevalenza di uno di questi due effetti sull'altro assume un'importanza notevole nella costruzione dell'offerta di lavoro, andiamo pertanto ad analizzarli maggiormente in dettaglio.

Partiamo da un vincolo con inclinazione W (salario inteso in termini reali) (graf.10), il punto di ottimo sarà E ; se il salario aumenta a $W2$ (salario reale), il vincolo ruota facendo perno sul punto di intersezione con l'asse verticale ed il nuovo punto di ottimo sarà E' .

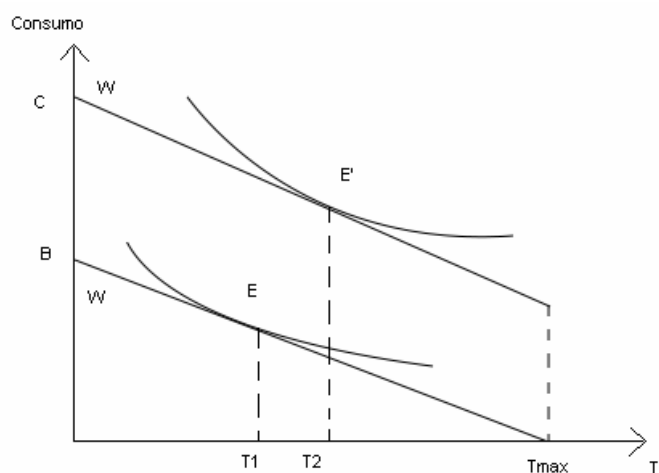
Se prevale l'effetto reddito avremo la situazione descritta nel graf.10



la correlazione tra il salario e le ore dedicate al lavoro sarà negativa. Se, invece, prevale l'effetto sostituzione (graf.11) avremo una situazione di questo tipo



Se, invece è il reddito non da lavoro ad aumentare, avremo una situazione di questo tipo

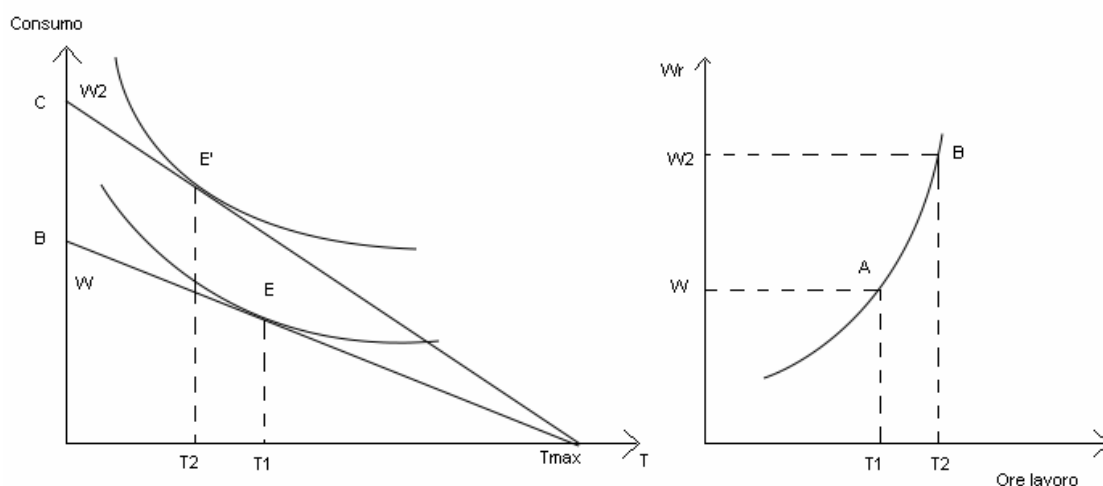


graf.12

quindi, supponendo la normalità del bene “tempo libero” risulta evidente che il solo effetto è l’effetto reddito.

A priori non è possibile stabilire quale dei due effetti prevarrà, anche se, generalmente è possibile dire che nei soggetti con orari di lavoro lunghi prevale l’effetto reddito, mentre prevale l’altro effetto per coloro che hanno poche ore lavorative¹¹.

Per quanto riguarda l’offerta individuale è possibile rappresentarla sul piano salario – orario lavorativo; nel caso prevalga l’effetto sostituzione avremo

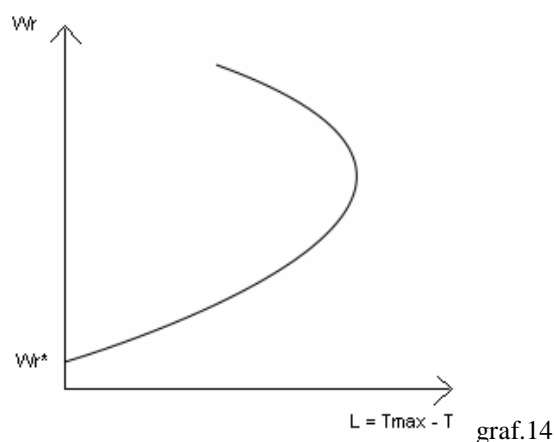


graf.13

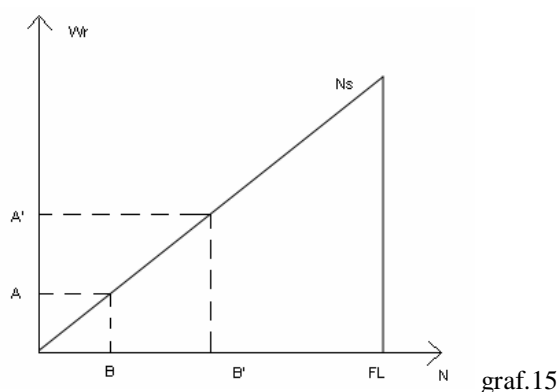
Partendo dall’offerta individuale possiamo, inoltre, dire che l’offerta aumenta all’aumentare del salario reale e l’inclinazione positiva si avrà solo fino a quando

¹¹ Ibid, p 65

l'effetto sostituzione è maggiore rispetto all'effetto reddito: ad un livello molto alto di salari, potrà prevalere l'effetto reddito e l'offerta di lavoro “curva all'indietro”¹²



Per costruire la funzione di offerta di lavoro di mercato, occorre andare a vedere, data la forza lavoro, quale è il numero di soggetti disposti a lavorare per ogni livello di salario reale. Alla luce di questo, è fondamentale che i salari reali “minimi” (a partire dai quali sono disposti a lavorare) per ciascun lavoratore siano diversi, quantomeno per classi, dagli altri, altrimenti la funzione di offerta sarebbe orizzontale. La tipica rappresentazione della curva d'offerta “aggregata” di lavoro è la seguente



L'offerta di lavoro di mercato è data dalla somma delle curve individua, che, come visto, possono essere inclinata negativamente, allora è lecito domandarsi il motivo per cui l'offerta di mercato è generalmente rappresentata positivamente inclinata. La tipica rappresentazione dell'offerta di mercato si basa sulla

¹² *Ibid*, p 66

semplice considerazione che a seguito di un aumento salariale, oltre agli effetti dovuti ai lavoratori che già operano sul mercato, vi saranno anche quelli di nuovi soggetti, difatti per ogni livello di reddito, un incremento salariale attira nuovi lavoratori nel mercato del lavoro accrescendo conseguentemente l'offerta complessiva¹³.

1.1.3 L'equilibrio di mercato

Le ipotesi fondamentali della perfetta concorrenzialità nel mercato del lavoro sono:

- perfetta informazione (completa e simmetrica): viene ipotizzata la totale assenza di imperfezioni informative
- completa omogeneità degli input (in particolare il lavoro) e dell'output (beni e servizi)
- costi di entrata e uscita dal mercato nulli
- assenza di frizioni nell'incontro tra domanda e offerta
- il numero di lavoratori e imprese è elevato ed entrambe considerano i prezzi, rispettivamente dei prodotti e del lavoro, esogeni.

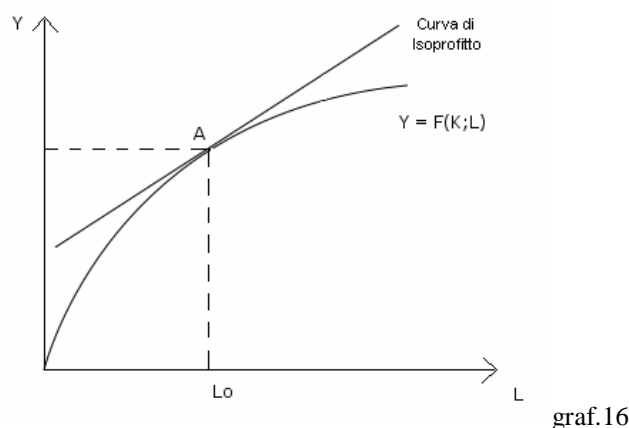
La perfetta concorrenzialità implica che le dinamiche tra i soggetti che operano all'interno di un mercato esauriscono tutte le possibili fonti di guadagno: se infatti ci fossero delle imprese con profitti positivi, questo indurrebbe altre imprese ad entrare nel mercato e se il numero di imprese tende ad infinito, allora i profitti si annullano¹⁴. In concorrenza perfetta, vale, inoltre, la legge del prezzo unico in base alla quale prodotti omogenei in equilibrio devono avere lo stesso prezzo.

¹³ Dornbusch R., Fischer S., *Macroeconomia*, il Mulino, p 279. 1995

¹⁴ Via via che le imprese entrano nel mercato spinte dalla possibilità di guadagni positivi, vi sarà un aumento dell'offerta di beni che causerà la riduzione dei prezzi con il conseguente assottigliamento dei profitti.

Abbiamo fino a questo punto visto come derivare concettualmente la domanda e l'offerta in ambito neoclassico sotto l'ipotesi di perfetta concorrenzialità e assenza di attriti: l'equilibrio è dato dall'intersezione tra la domanda e l'offerta e l'ipotesi di perfetta flessibilità dei prezzi garantisce di stare sempre al pieno impiego.

Abbiamo già visto che ciascuna impresa ha come obiettivo la massimizzazione dei profitti, rispetto a L dato che siamo nel breve periodo, e il vincolo è rappresentato dalla “realità tecnologica” sintetizzata nella funzione di produzione; la situazione di ottimo si ha pertanto nel punto di tangenza tra la curva di isoprofitto più alta e la funzione di produzione¹⁵



La pendenza della curva di isoprofitto è il salario reale¹⁶, mentre la pendenza della funzione di produzione è la produttività marginale, nel punto A si quindi la massimizzazione dei profitti¹⁷.

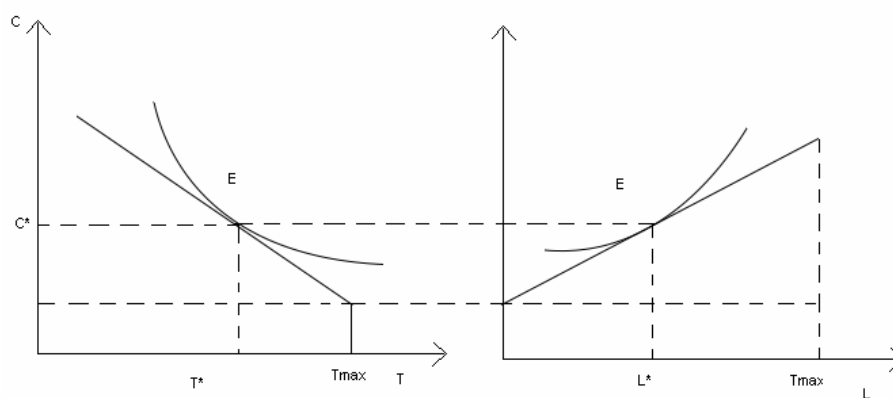
Resta da analizzare la fonte dei redditi non da lavoro e la destinazione dei profitti dell'impresa. Preliminarmente possiamo rappresentare (graf.17) il vincolo di bilancio del lavoratore e le sue curve di indifferenza in funzione non più del tempo libero ma del tempo dedicato al lavoro¹⁸, graficamente

¹⁵ Brucchi L., *op.cit.*, p 110

¹⁶ In quanto lo scopo è la massimizzazione rispetto L di $PY-WL-RK$, sotto il vincolo della funzione di produzione

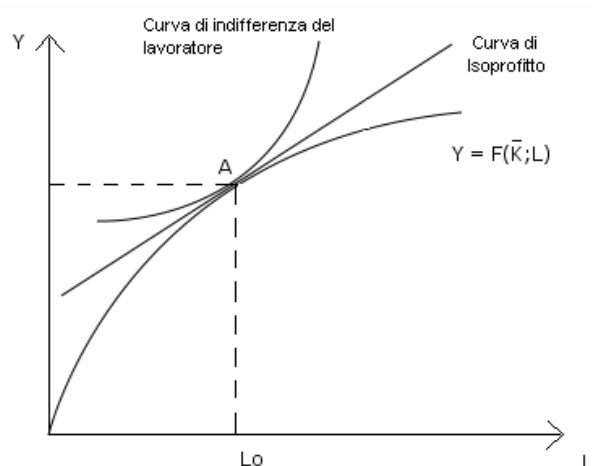
¹⁷ Se ci fosse un aumento del salario reale, la curva di isoprofitto sarebbe più inclinata e il punto di ottimo sarebbe più “in basso“, lungo la funzione di produzione, rispetto ad A e quindi l'impresa domanderebbe meno lavoro e offrirebbe un livello di produzione inferiore.

¹⁸ *Ibid.*, p 109



graf.17

Ipotizzando che le imprese siano di proprietà dei lavoratori, i redditi non da lavoro e i profitti d'impresa coincideranno, data questa ipotesi è possibile ricavare la seguente rappresentazione grafica¹⁹.



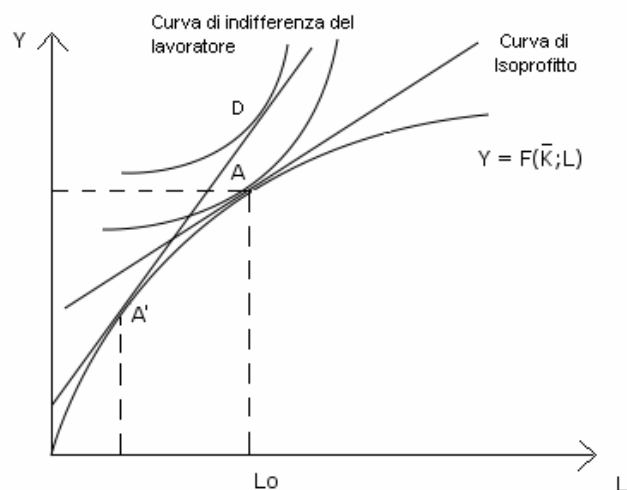
graf.18

Nel punto A, il lavoro offerto corrisponde esattamente al livello domandato e quindi non c'è disoccupazione involontaria, similmente il livello di prodotto offerto dalle imprese coincide esattamente con quello domandato; il reddito non da lavoro (OB) (graf.18) coincide con i profitti e il reddito da lavoro (BC) corrisponde con il costo del lavoro sopportato dalle imprese.

Supponiamo di trovarci in una situazione in cui il salario è maggiore (per motivi esogeni, come ad esempio motivi istituzionali) rispetto al livello d'equilibrio²⁰

¹⁹ *Ibid.*, p 111

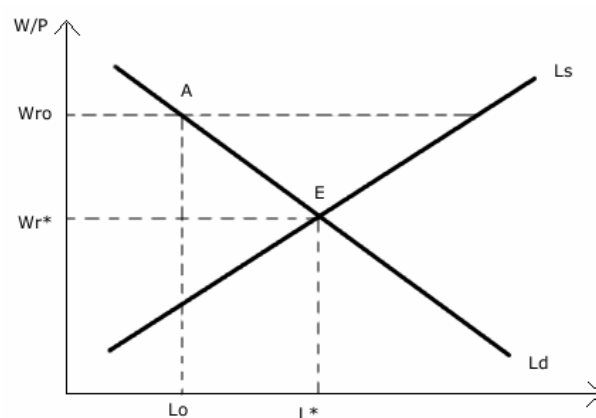
²⁰ *Ibid.*, p 114



graf.19

quindi vi sarà un aumento dell'offerta di lavoro (punto D) (graf.19) ma congiuntamente anche una riduzione della domanda di lavoro (punto A') e quindi è necessario che la retribuzione reale reagisca prontamente (istantaneamente) per garantire l'uguaglianza tra domanda e offerta, altrimenti si verrebbe ad avere una situazione di disoccupazione involontaria. I neoclassici ipotizzano aggiustamenti istantanei e pertanto la disoccupazione involontaria non trova spazio poiché verrebbe immediatamente "riassorbita" per effetto della riduzione del salario reale.

Riprendiamo a questo punto il grafico (graf. 1) introdotti all'inizio del capitolo

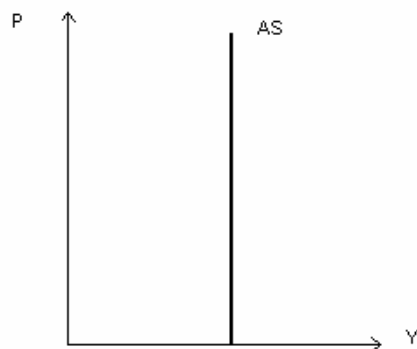


al livello del salario W_{ro} le imprese, necessitando di una produttività marginale superiore per compensare i maggiori costi, domanderanno meno lavoro e i lavoratori aumenteranno l'offerta: la pressione competitiva tra i lavoratori,

ridurrà il salario reale e pertanto avremo che l'impresa aumenterà la domanda in quanto il costo del lavoro sta diminuendo e questo meccanismo impedisce la creazione e la persistenza di disoccupazione involontaria facendo giungere all'equilibrio²¹.

I prezzi si aggiustano in modo immediato in relazione all'eccesso di offerta o di domanda e quindi riescono sempre in modo istantaneo ad aggiustare gli eccessi nel mercato del lavoro. Ipotesi fondamentale è che tutti i lavoratori attribuiscono lo stesso valore al tempo libero e che la retribuzione che ciascuna impresa paga nel mercato sia la stessa.

A questo punto è necessario sottolineare che l'ipotesi di assenza di attriti implica una curva di offerta aggregata verticale, cioè perfettamente anelastica²² come rappresentata dal seguente grafico



graf.20

il livello di prodotto offerto al pieno impiego è offerto a qualsiasi livello di prezzo, quindi se i prezzi aumentano aumenterà anche il salario nominale (rispondendo immediatamente) e quindi non avremo variazioni reali.

Il mercato del lavoro, quindi, è necessario per consentire a tutti gli altri mercati (capitali e beni) di giungere all'equilibrio di pieno impiego: dato il livello di

²¹ La flessibilità del salario monetario è sufficiente ma non necessaria per giungere nuovamente all'equilibrio: l'aumento dei prezzi dovuto all'eccesso di domanda nel mercato di beni, infatti, anche a parità di salario monetario, determina una riduzione del salario reale.

²² Dornbusch R., *op. cit.*, pg 280

occupazione, si ricava il livello di produzione che le imprese offrono; per avere l'equilibrio nel mercato dei beni, occorre che la domanda aggregata eguagli l'offerta e la variabile che gioca questo ruolo è il tasso di interesse reale determinato sul mercato dei capitali.

L'impianto neoclassico si completa, consentendo di giungere ai tipici risultati del modello, introducendo e integrando la così detta "legge degli sbocchi": Say sosteneva che la "volontà di consumo" fosse limitata solo dalla disponibilità di mezzi monetari che trovavano origine dalla vendita di altri prodotti o dai redditi, quindi "se è con i prodotti che si comprano i prodotti" ogni merce verrà acquistata quanto più aumenta la produzione.

I neoclassici, in sostanza, ritenevano che fosse l'offerta a creare la propria domanda; viene negata in questo modo la possibilità di crisi di sovrapproduzione generale.

1.2 Osservazioni generali e critiche sul modello neoclassico, cenni sull'approccio keynesiano

Alcuni tra i principali risultati della scuola neoclassica, derivano dall'utilizzo di uno strumento teorico molto utile, la statica comparata.

Usare questo strumento, come abbiamo visto anche in precedenza, è facile e consente con semplicità di capire alcune questioni che stanno dietro a modelli che altrimenti sembrerebbero fine a se stessi; seppur di facile utilizzo, la statica comparata non può tuttavia integrare la dimensione temporale. L'analisi statica cristallizza una certa situazione, fissando tutti i parametri tranne quello oggetto d'analisi: in questo modo si riesce a capire l'interazione "base" del parametro con gli altri, ma non si considera minimamente un'interazione a livello più alto tra i parametri stessi. In sostanza, l'analisi neoclassica è limitativa poiché atemporale.

Un altro aspetto da sottolineare è l'ipotesi neoclassica di esogeneità dell'efficienza d'impresa: la tesi secondo la quale la concorrenzialità del mercato induce l'impresa a minimizzare i costi, consentendole di operare in modo efficiente, è condivisa da moltissimi studiosi ed è anche supportata da numerose verifiche empiriche; questo aspetto, tuttavia, non è minimamente affrontato in ambito neoclassico nel quale l'efficienza viene semplicemente ipotizzata. Risulta plausibile, invece, ipotizzare che il grado di efficienza di un'impresa debba essere considerato come endogeno²³; l'analisi neoclassica, però, manca di qualsiasi modello relativo alle determinanti dell'efficienza d'impresa²⁴.

²³ Graffi A., Il ruolo dell'informazione nella relazione tra concorrenza ed inefficienza-x: ambiguità teoriche della "product market competition", Liuc Papers n. 197, Serie Economia e Impresa, 51, p. 1. 2006

²⁴ Uno dei primi studi teorici, finalizzati a dimostrare la positiva correlazione tra concorrenza e efficienza è stato Leibenstein (1966) la cui impostazione si basava sull'incompletezza dei contratti: l'impresa sostiene dei costi per monitorare i lavoratori ed evitare che siano "inattivi", il mancato impegno genera inefficienza interna che impedisce la minimizzazione dei costi, ma la concorrenza esterna riduce l'inefficienza interna quindi l'efficienza è tanto maggiore quanto maggiore è il grado di concorrenzialità. L'inefficienza interna nasce dall'assenza di adeguati incentivi all'efficienza organizzativa e lavorativa.

E' opportuno porre l'attenzione anche sul graduale cambiamento nell'approccio metodologico che ha spostato l'attenzione dall'analisi della crescita economica allo studio delle proprietà degli equilibri, tipicamente walrasiani. La matematicizzazione dell'economia, introdotta dai neoclassici, ha consentito di rappresentare in modo formale e accurato aspetti della realtà prima sconosciuti, ma ha anche "limitato" l'analisi teorica facendola divenire un processo virtuale dipendente da deduzioni basate su un sistema formale di ipotesi o assiomi assunti ad hoc.

L'approccio neoclassico, anche se molto "potente" in termini descrittivi, rischia di essere un mero sforzo di virtuosismo tecnico: il sistema neoclassico, in realtà, sembrerebbe basato su un impianto assiomatico non dimostrabile e per questo sembrerebbe perdere la sua capacità esplicativa, andando a studiare "pure" relazioni prezzi – quantità.

Un altro punto suscettibile di critica è rappresentato dall'idea dei soggetti economici ed in particolare dall'ipotesi di razionalità. La razionalità, fondamentale nella costruzione neoclassica, non sembra caratterizzare tutte le relazioni economiche, basti pensare ad esempio ai consumi. Inoltre, come detto in precedenza, un elemento essenziale per comprendere l'economia nella sua dinamica è la dimensione temporale che genera un problema enorme relativo all'incertezza sul futuro.

Gli agenti ignorano il futuro e questo genera enormi conseguenze sui comportamenti degli operatori, comportamenti che diventano adattivi, che si basano, cioè, sul passato. L'incertezza però, oltre che relativa al futuro, si ha anche nel presente: i risultati neoclassici si basano sull'ipotesi che tutti i soggetti sappiamo "tutto", che ci sia cioè una perfetta simmetria informativa tra tutte le parti che non consenta comportamenti opportunistici.

Nella realtà le asimmetrie informative pervadono tutti i mercati e generano, come vedremo, enormi problemi.

Un'ulteriore limite del costrutto neoclassico è la visione delle istituzioni, viste non come elementi dinamici ma piuttosto come strumenti analitici d'aggiustamento: il contratto, di fatto, non assume rilevanza in un contesto neoclassico; il “banditore” in realtà è una figura posta ad hoc di cui si può fare del tutto a meno; il mercato non è altro che il risultato delle azioni di soggetti “matematici” e l'insieme dei mercati costituisce il sistema economico; ma ai mercati neoclassici non servono “regole” in quanto rappresentano relazioni virtuali tra soggetti fuori dalla dimensione temporale.

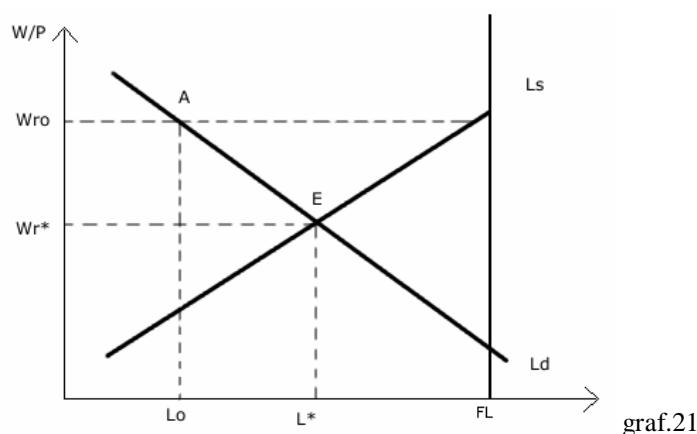
Da un punto di vista sociale è opportuno sottolineare anche che l'impianto neoclassico fornisce, come già intuito dalle considerazioni precedenti, una visione riduttiva di molti aspetti della realtà:

- i bisogni umani sono “solo” quelli che possono essere soddisfatti grazie al mercato e tutti gli agenti sono egoisti e non vi è lo spazio per considerazioni etiche
- la libertà consiste nello scegliere beni di consumo e la società è la “semplice” somma degli individui
- l'obiettivo è il solo guadagno e la “pura” logica del profitto non conosce limiti e regole etiche

Partendo da queste ultime considerazioni di carattere sociale è appropriato notare che a differenza degli economisti neoclassici, Keynes, la cui analisi segna il punto di rottura dal paradigma neoclassico, non riuscì (forse volutamente) a scindere l'analisi teorica dalla propria “filosofia morale” giungendo addirittura ad auspicarsi “una profonda revisione” del capitalismo stesso che ne correggesse i difetti preservandone tuttavia i caratteri fondamentali”²⁵; l'economista inglese, infatti, denuncia l'intrinseca instabilità del sistema capitalistico in assenza di automatismi di correzione.

²⁵ Pavanelli G., *op. cit.*, p 323

In sostanza, Keynes critica parte dell'impianto neoclassico, proponendo un nuovo approccio, basato, però, anche su "vecchie idee"; riprendiamo la tipica rappresentazione del mercato del lavoro



il numero di lavoratori totali in grado di lavorare è dato da FL che indica la forza lavoro, per i neoclassici al livello L^* si ha piena occupazione e i lavoratori rappresentati dalla differenza tra la forza lavoro e i lavoratori d'equilibrio rappresentano disoccupati volontari; viene esclusa qualsiasi tipo di disoccupazione involontaria.

Keynes accettò il primo postulato neoclassico, che prevede l'uguaglianza tra il salario reale e la produttività marginale, criticando il secondo che prevede invece l'uguaglianza tra il salario reale e la disutilità marginale del lavoro:

- la contrazione salariale è relativa al salario nominale, non a quello reale: i lavoratori si oppongono con forza ad una riduzione del salario nominale, mentre una riduzione di quello reale è, almeno inizialmente, non accompagnata da rivendicazioni
- i lavoratori non sarebbero comunque in grado di individuare il livello del salario reale
- una riduzione del salario nominale non necessariamente si traduce in una riduzione della stessa entità del salario reale, ipotesi implicita nella costruzione neoclassica ma in realtà in contrasto con le altre assunzioni:

secondo i neoclassici il prezzo tende ad eguagliare il costo marginale che dipende dal salario nominale, quindi una riduzione del salari nominale si traduce in una riduzione dei prezzi lasciando (quasi) invariato il salario reale.

In sostanza, per Keynes il secondo postulato va rigettato e l'occupazione, nell'ambito del mercato del lavoro, va fissata esogenamente sulla base della domanda aggregata.

Il rifiuto di Keynes del secondo postulato (rifiuto che implica la possibilità di disoccupazione involontaria) non si estende anche al primo, quindi per Keynes, a differenza degli economisti neoclassici, l'equilibrio di mercato è sempre, e solo, lungo la curva di domanda e quindi, possibilmente, anche al di sotto del pieno impiego.

E' opportuno chiarire più in dettaglio questo aspetto: per Keynes le cause della disoccupazione non vanno cercate all'interno del mercato del lavoro; secondo l'economista inglese, infatti, la situazione del mercato del lavoro dipende in gran misura dal mercato dei beni. Keynes pone particolare attenzione sulla domanda aggregata (in particolare sui consumi e investimenti) e sul prodotto nazionale; se non vi è nessuna variazione della domanda aggregata, nessuna dinamica interna al mercato del lavoro è in grado di modificare il livello d'occupazione.

Keynes pone, quindi, la domanda aggregata al centro della sua analisi teorica creando in questo modo una netta contrapposizione con l'impostazione neoclassica della "legge di Say": se le imprese non vendono tutta la produzione, si viene ad avere disoccupazione da insufficienza della domanda, si viene ad avere, cioè, una situazione nella quale i lavoratori sarebbero disposti a lavorare al salario corrente, ma questo non è possibile per tutti i lavoratori a causa della riduzione della domanda di lavoro dovuta alla riduzione della produzione a seguito del calo della domanda aggregata.

Successivamente a questa dinamica si ha una riduzione del livello di attività economica e pertanto il sistema economico può trovarsi in una situazione

d'equilibrio di sottoccupazione dalla quale non esistono dei meccanismi automatici per giungere al pieno impiego: l'unico strumento adatto è la politica economica a sostegno della domanda aggregata.

Ricapitolando, se la domanda aggregata è scarsa, le imprese riducono l'offerta fino ad eguagliare la domanda e pertanto esistono situazioni in cui il livello di attività economica è inferiore di quello potenziale; l'offerta aggregata può essere rappresentata come una retta orizzontale in quanto, dato l'eccesso di capacità produttiva, le imprese possono offrire, per un dato livello di prezzo, qualsiasi livello produttivo. Se la domanda cala, anche l'offerta diminuisce e diminuisce pertanto anche il numero di lavoratori impiegati e questo genera disoccupazione che persiste poiché i salari non possono scendere sotto un certo livello, si ha quindi un equilibrio di sottoccupazione.

Per concludere, Keynes sottolinea che una riduzione del salario nominale può avere effetti sull'occupazione solo se questa riduzione determina l'aumento di una delle componenti della domanda aggregata; Keynes, però, sottolinea che una riduzione del salario nominale comporta una riduzione del costo marginale di produzione e quindi un aumento dell'offerta di beni ma questo non implica necessariamente un aumento della domanda aggregata: se la domanda aggregata non aumenta si viene ad avere solo una riduzione dei prezzi, riduzione che lascia immutato il salario reale e quindi anche la domanda e l'offerta di lavoro; secondo Keynes, il salario reale potrebbe addirittura aumentare a seguito di una riduzione del salario nominale determinata da un calo della domanda aggregata.

Per i neoclassici la relazione causa – effetto è semplice, ossia una riduzione (aumento) del salario nominale comporta una riduzione (aumento) del salario reale e quindi un aumento (diminuzione) dell'occupazione.

Per Keynes, invece, una variazione del salario nominale determina una variazione, i cui esiti sono incerti, della domanda aggregata; ad esempio una riduzione del salario nominale può determinare un calo della domanda aggregata e successivamente una diminuzione dell'occupazione: Keynes, però, accetta il

primo postulato neoclassico e pertanto un calo della domanda aggregata comporta necessariamente un aumento del salario reale, dovuto alla diminuzione dei prezzi.

Occorre, quindi, domandarsi quali siano le cause che, secondo Keynes, determinano un calo della domanda aggregata a seguito di una riduzione dei salari nominali, occorre analizzare, cioè, le variabili da cui dipende la domanda stessa come ad esempio la propensione aggregata al consumo, l'efficienza marginale del capitale ed il tasso di interesse.

Una riduzione dei salari nominali, e quindi dei costi di produzione, si traduce in un calo dei prezzi che genera effetti redistributivi a favore di coloro che ricevono un reddito fisso determinando, presumibilmente, un calo della propensione al consumo²⁶.

Possiamo, inoltre, osservare che un calo dei costi di produzione, e successivamente dei prezzi, dovuto alla riduzione del salario nominale determina un aumento in termini reali dell'ammontare dei debiti scoraggiando quindi gli investimenti e portando ad effetti depressivi.

In conclusione, per Keynes, la stimolazione della domanda aggregata, strumento per agire sull'occupazione, deve avvenire tramite strumenti finalizzati a tale scopo e non in modo indiretto.

1.2.1 L'ipotesi di perfetta simmetria informativa e i limiti della "transazione di mercato"

L'intera impostazione neoclassica e i risultati che essa implica si basano, come parzialmente visto, in gran misura sull'ipotesi di perfetta simmetria delle informazioni e sulla presenza di istituzioni "virtuali" quali il "contratto", il "banditore" ed il "mercato".

²⁶ *Ibid.*, p 322

Per quanto riguarda più specificatamente il contratto in ambito neoclassico, occorre sottolineare che il suo ruolo trova origine e si esaurisce nella semplice individuazione della quantità dei beni o servizi scambiati e dei prezzi: il contratto, inteso sempre come completo ed esplicito, viene quindi visto non come uno strumento per l'attuazione forzata dello scambio, ma come l'oggetto economico che definisce lo scambio in sé.

Questo è il motivo per cui il ruolo del contratto in ambito neoclassico è puramente virtuale anche se in sua assenza, di fatto, non è possibile pensare ad alcuna transazione economica.

Così come il contratto, il ruolo del “banditore” non è altro che uno strumento “ideale” finalizzato alla convergenza all'equilibrio; similmente il “mercato” assume connotati astratti trovando la propria ragion d'essere nella rappresentazione delle relazioni tra soggetti teorici al di fuori di una concreta dimensione reale.

Quanto fino ad ora detto si sintetizza nell'espressione “transazione di mercato ideale”²⁷ che si contrappone alla “transazione relazionale” propria, invece, della teoria dei contratti.

Per gli economisti neoclassici è del tutto irrilevante l'identità delle parti coinvolte, inoltre è ipotizzata l'omogeneità di qualsiasi fattore scambiato all'interno del mercato: tutti i lavoratori sono uguali così come tutte le merci e servizi. Un altro aspetto fondamentale è la totale assenza di incertezza all'interno del mercato. Accanto a queste assunzioni implicite ve ne sono altre che forse giocano un maggior contributo:

- i termini contrattuali sono “perfetti” e consentono, in qualsiasi circostanza, l'intervento di una giudice (o corte) per imporne l'attuazione,

²⁷ Ruiz M. L., *Mercato o contratto?*, Stato o mercato? Intervento pubblico e architettura dei mercati, p. 167, Pavia, vol. I. 2001

ma di fatto la possibilità di ricorrere in modo “perfetto” ad una corte, evita a priori il ricorso stesso poiché, data questa consapevolezza, nessuno verrà meno agli accordi presi

- l'unico costo della transazione è quella per la transazione in sé; la stesura del contratto non presenta costi così come la raccolta di tutte le informazioni necessarie per decidere se realizzare o meno lo scambio.

Un aspetto fondamentale per il completamento della transazione è, infatti, la “conoscenza” dell'oggetto scambiato. Ovviamente, agendo su un mercato, si intende, ad esempio, comprare una certa merce che corrisponde a certe nostre esigenze, ma questo presuppone la conoscenza delle caratteristiche del bene che si intende comprare.

Nell'impianto neoclassico l'attività di raccolta delle informazioni può essere fatta in modo completo e gratuito: se un soggetto, ad esempio, desidera acquistare un prodotto elettronico, l'impostazione neoclassica suggerisce che l'acquirente possa, in modo del tutto privo di costi, informarsi in modo completo circa tutti gli aspetti del prodotto e prendere una decisione condizionata a tali informazioni. Similmente, un datore di lavoro, in ambito neoclassico, conosce perfettamente tutti gli aspetti relativi ai propri dipendenti ed in particolare le loro *performance*. Secondo i neoclassici la completezza e la simmetria delle informazioni è del tutto fondamentale oltre che dal punto di vista prettamente analitico anche sul piano concettuale.

La completezza si rende necessaria alla luce della razionalità delle scelte dei soggetti: i soggetti non potrebbero agire in modo razionale se non conoscessero tutti gli aspetti della contrattazione e quindi non potrebbero massimizzare i propri obiettivi (nucleo centrale dell'analisi neoclassica). La simmetria, invece, è fondamentale per evitare situazioni di potere di mercato e, come vedremo, rischi di opportunismo.

Come risulta evidente, ipotesi di questo tipo trovano la propria ragion d'essere solo all'interno di un mondo "virtuale" (come è l'impianto neoclassico) dove per altro non necessitano neanche di essere esplicitate in quanto tutti i prodotti sono identici e quindi l'attività di raccolta di informazioni è ulteriormente semplificata poiché va fatta per un solo prodotto per ogni mercato.

Data la perfetta simmetria informativa, costi di transazione nulli e una totale certezza all'interno dei mercati, è ovvio come sia possibile giungere, con facilità, ad un prezzo e ad una quantità d'equilibrio non influenzabili dai soggetti (intesi in termini atomistici), prezzo efficiente che racchiude in sé l'intera dimensione informativa necessaria per comprendere tutti gli aspetti della contrattazione.

Per completezza occorre chiarire che l'eventuale mancanza di "certezza" non è né necessaria né sufficiente per poter dire che la "transazione di mercato" non può essere realizzata: l'incertezza non è necessaria perché nel momento in cui ci sono costi specifici è molto probabile che la forma di scambio si configuri come "transazione relazionale" anche nel caso di certezza; l'incertezza non è neppure una condizione sufficiente per essere costretti ad abbandonare la "relazione di mercato", i mercati contingenti alla Arrow – Debreu sono stati creati apposta per considerare tutte le possibili situazioni future in un contesto di incertezza e permettono di giungere ad una situazione efficiente in senso paretiano.²⁸

Un aspetto rilevante che necessita di essere nuovamente sottolineato è la totale indifferenza dell'identità delle parti; in ambito neoclassico un contratto può essere ceduto, senza costi di nessun tipo, ad un altro soggetto (del tutto identico al precedente), una relazione "fiduciaria" è del tutto priva di significato in questo ambito oltre che ad essere del tutto inutile.

Nella realtà è difficile riscontrare situazioni che corrispondono (o almeno tendono) all'impostazione neoclassica: in ambito finanziario, ad esempio, è

²⁸ *Ibid.*, p 170

possibile individuare dei tipi di contrattazioni che si avvicinano all'impostazione neoclassica grazie all'elevato numero di operatori che fa tendere ad una situazione di concorrenza perfetta, tuttavia anche i mercati finanziari sono fortemente soggetti al problema delle asimmetrie informative (condizione che ci distanzia dal paradigma neoclassico); un esempio significativo è rappresentato dalle contrattazioni spot dove sono disponibili al momento dello scambio tutte le informazioni rilevanti (in quanto si presuppone che per periodi vicini i costi di informazione siano molto bassi) ²⁹ e dove lo scambio delle obbligazioni tra le parti è simultaneo, e conclusivo.

Situazioni di questo tipo nella realtà, però, oltre ad essere rare sono anche molto limitative perché devono essere necessariamente simultanee e quindi non consentono la reiterazione delle transazione, alla base invece, della maggior parte dei mercati come ad esempio il mercato del lavoro.

Nella realtà si è visto, invece, come sia assolutamente rilevante l'identità (e quindi la possibilità di distinguerli) dei soggetti; un po' alla volta è possibile intaccare le ipotesi base dell'assunto neoclassico cominciando, ad esempio, ad osservare che l'ipotesi di costi nulli di raccolta delle informazioni è del tutto irrealistica, alla luce del fatto che vi è una pluralità di beni concorrenti ma non omogenei.

Ad esempio possiamo osservare che la relazione lavorativa, presumibilmente reiterabile e reiterata anche in ambito neoclassico, non sia priva di costi di cessazione, dovuti ad esempio ad investimenti specifici, come invece previsto dall'impianto neoclassico.

²⁹ *Ibid.*, p 186

2. Asimmetrie informative: l'economia dei contratti. L'approccio dei contratti d'agenzia e il problema del rischio.

Come visto nel capitolo precedente, un significativo limite dell'impostazione neoclassica è rappresentato dall'ipotesi di perfetta simmetria informativa; tale ipotesi, come già sottolineato, è difficilmente riscontrabile nella realtà, dove soggetti diversi hanno, tipicamente, accesso a informazioni diverse e quindi si assiste ad una forte diversificazione distributiva dell'informazione.

Anche nell'analisi teorica, pertanto, dobbiamo affrontare la realtà dei fatti ed accettare che i soggetti interessati ad uno scambio non possiedono e, tendenzialmente, non sono in grado di possedere¹ tutte le informazioni rilevanti circa la controparte o circa la natura e le caratteristiche del prodotto oggetto dello scambio. Prendere atto della realtà, però, non è sufficiente: per capire i principali aspetti del problema e analizzarne le principali implicazioni risulta necessario definire in modo formale il problema stesso. E' stata questa esigenza a dar vita alla teoria dei contratti.

Un altro importante aspetto molto rilevante nella realtà è il problema dell'incertezza che rende necessario l'utilizzo di approcci differenti da quello "classico" della domanda e offerta: l'incertezza ed il rischio sono elementi indissolubili da qualsiasi realtà economica e fanno sì che le scelte dei soggetti siano sempre caratterizzate da un certo grado di aleatorietà, in quanto una serie di eventi indipendenti dalle azioni del soggetto, possono condizionare il risultato che ci si attende.

Il problema dell'incertezza è stato affrontato in diversi contesti teorici; ad esempio nel mercato dei beni contingenti alla Arrow – Debreu, pur essendoci

¹ Virtualmente sarebbe possibile ottenere informazioni in modo "perfetto", l'impossibilità di fatto deriva dall'ipotesi realistica che per poter fare ciò è necessario sostenere ingenti costi.

incertezza² e pur essendo il momento della contrattazione diverso da quello dello scambio e del pagamento, viene definito a priori il comportamento per ogni possibile stato futuro del mondo.

Sebbene non si conosca con precisione quale sarà l'evento che si realizzerà, risulta, tuttavia, possibile costruire per ogni bene (anche quelli futuri) un mercato completo. Nella realtà, un'operazione di pianificazione contrattuale alla Arrow – Debreu è irrealizzabile in quanto sorgono problemi legati sia all'incompletezza contrattuale (aspetto che verrà maggiormente approfondito nel capitolo seguente) sia alla struttura e al funzionamento del mercato stesso.

Come vedremo immediatamente, il problema dell'incertezza gioca un ruolo centrale nel modello d'agenzia, introdotto nel paragrafo successivo; in particolare introducendo l'incertezza, gli operatori non conoscono le possibili situazioni future e quindi non è possibile individuare con certezza tutte le future grandezze economiche, né, a partire da queste, procedere a ritroso per analizzare le azioni che le hanno prodotte: questa è la tipica situazione di qualsiasi mercato.

Il principale oggetto d'analisi in questo lavoro, però, è rappresentato dalle asimmetrie informative, vero nucleo della teoria dei contratti, anche se ovviamente è impossibile prescindere dal problema dell'incertezza. Andiamo a vedere, pertanto, cosa si intende per asimmetria informativa: in generale, la creazione di un mercato presenta enormi costi di transazione, inoltre sono molteplici le cause di fallimento dei mercati, una delle quali è appunto l'asimmetria informativa. Procediamo per gradi e definiamo in primo luogo cosa si intende per fallimento del mercato.

Il primo teorema dell'economia del benessere stabilisce che ogni economia perfettamente competitiva raggiunge e mantiene una situazione ottimale in senso paretiano.

² Il concetto di incertezza viene fatto, tendenzialmente, coincidere con quello di rischio, tuttavia alcuni studiosi hanno operato una distinzione. Ad esempio, Knight per rischio intende che sia possibile individuare gli eventi futuri e attribuirgli una certa probabilità di realizzazione, mentre per incertezza intende una situazione in cui non sia possibile individuare neanche le possibili situazioni future; in questa ottica, il modello di Arrow e Debreu non presenta incertezza.

Come già diffusamente trattato nel capitolo precedente, l'impostazione neoclassica consente di raggiungere una situazione di equilibrio grazie al fatto che né il prezzo dell'output né quello degli input sono modificabili dai singoli soggetti; le dinamiche di mercato, tramite il meccanismo dei prezzi, quindi, consentono, a partire da qualsiasi situazione di squilibrio, di ottenere un equilibrio efficiente in termini paretiani che comporta l'uguaglianza al margine tra il beneficio ed il costo sociale derivanti dalla transazione.

Si parla di fallimento del mercato, quindi, ogni qual volta ci si distanzia dalla situazione precedentemente descritta; la mancanza di informazione completa e simmetrica comporta necessariamente l'allontanamento da una situazione ottima in senso paretiano (quella che gli economisti definiscono *first-best*) e può determinare la fine del mercato stesso.

Per chiarire questo punto, consideriamo una semplificazione del mercato delle auto usate (mercato dei "bidoni") introdotto da Akerlof³, dove i venditori conoscono la qualità dell'auto, mentre gli acquirenti la ignorano e pertanto, dato l'opportunismo delle parti, vi è il concreto pericolo di prendere "un bidone".

Supponiamo che vi siano N auto (metà efficienti, e metà inefficienti) e che i proprietari siano disposti a vendere⁴ le automobili efficienti ed inefficienti, rispettivamente a $P_e = 6$ e $P_i = 3$, mentre gli acquirenti siano disposti a comprare le auto efficienti e i "bidoni" rispettivamente a $\underline{P_e} = 6,2$ e $\underline{P_i} = 3,2$. In presenza di simmetria informativa, tutte le N auto verranno scambiate⁵ e i prezzi di equilibrio P_e^* e P_i^* saranno contenuti, rispettivamente negli intervalli (6; 6,2) e (3; 3,2).

In condizione di asimmetria, invece, tutte le auto vengono offerte al prezzo P_e ; i proprietari tentano di trarre vantaggio dalla situazione, gli acquirenti, invece, conoscono solo la probabilità ($\text{prob}=0,5$) di prendere un "bidone" e sono disposti a pagare un'auto qualsiasi

$$V = 0,5 \underline{P_e} + 0,5 \underline{P_i} = 4,7$$

³ Akerlof A. G., *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, The Quarterly Journal of Economics, Vol. 84, No. 3, pp 488-500. 1970

⁴ Il prezzo (minimo) a cui sono disposti a vendere le auto è il valore che attribuiscono all'auto stessa.

⁵ Nell'ipotesi che vi sia un livello sufficiente di domanda.

e quindi risulta $V < P_e$, $V > P_i$. In questa situazione resteranno sul mercato solo i venditori di auto inefficienti (selezione avversa) e il volume degli scambi sarà inefficiente poiché i proprietari di auto in buono stato sarebbero disposti a vendere le automobili, ma ad un prezzo maggiore. In generale, sia γ la probabilità di comprare una buona auto, il prezzo che l'acquirente è disposto a pagare per un'auto qualsiasi sarà

$$V1 = \gamma \underline{P_e} + (1 - \gamma) \underline{P_i}$$

Se γ è sufficientemente alto, avremo $V1 > P_e$, quindi, nonostante l'asimmetria, tutte le auto vengono scambiate e il prezzo cade nell'intervallo $(P_e, V1)$ ⁶. Da un punto di vista analitico⁷, ponendo $V1 > P_e$, si ricava γ tale per cui tutte le N auto vengono vendute

$$\gamma > (P_e - \underline{P_i})(\underline{P_e} - \underline{P_i})^{-1}$$

Il volume degli scambi è uguale a quello che si avrebbe in simmetria, però i venditori di auto in cattivo stato realizzano maggiori profitti dei proprietari di auto buone che pertanto non hanno incentivo ad entrare nel mercato. Nel tempo, quindi, la qualità media delle auto usate in vendita è destinata a scendere.

In conclusione, i mercati dove ci sono asimmetrie informative sicuramente si distanzieranno da una situazione efficiente e potrebbero anche scomparire in favore di mercati dove c'è una relazione personale basata o sulla "fiducia" e/o su un eventuale costo che uno delle due parte accetta di accollarsi.

Il termine "fiducia" non è usato a caso: nel contesto di perfetta concorrenzialità, tutti i soggetti sono identici e possono essere sostituiti l'un con l'altro a costo nullo e pertanto è del tutto insignificante il così detto "problema della delega" infatti se un soggetto non è disposto a seguire le direttive, può essere sostituito ad un costo nullo. Il mondo economico, però, è caratterizzato da una fortissima

⁶ Similmente, se γ è sufficientemente basso, $V1$ risulterà inferiore rispetto al valore delle auto in cattivo stato, per cui il volume degli scambi sarebbe nullo.

⁷ Devono valere le seguenti condizioni $P_e > P_i$, $\underline{P_e} > \underline{P_i}$, $\underline{P_e} > P_e$, $\underline{P_i} > P_i$. In particolare, se non fossero verificate le ultime due ipotesi, non ci sarebbe, anche in condizione di simmetrie, nessuno scambio.

componente di individualità e le informazioni circa le modalità in cui un soggetto sceglie non sono disponibili in modo simmetrico e, quindi, il rapporto tra due soggetti economici può presentare dei problemi “impensabili” in ambito neoclassico. Risulta quindi chiaro il motivo per cui è stato usato il termine “fiducia”: nel mercato concorrenziale non serve (e non avrebbe neanche senso parlarne) una relazione di fiducia tra le parti; ciascuno può, senza costi, trovare un altro lavoratore (lavoro) muovendosi “istantaneamente” sul mercato; se, però, questo paradigma viene meno, risulta evidente che la presenza di costi, ad esempio di ricerca, porta a tentare di mantenere il rapporto economico e, quindi, la fiducia reciproca gioca un ruolo fondamentale in tal senso, diventando un vero e proprio bene.

In un’ottica economica, però, per fiducia non si intende “fede cieca” nei confronti di qualcuno: se un datore di lavoro ha fiducia (in senso economico) nei confronti di un dipendente, questo significa che grazie a degli incentivi, il datore di lavoro sa che il lavoratore si comporterà in modo corretto.

Un altro elemento che distanzia fortemente la nuova impostazione da quella neoclassica è, infatti, rappresentato dal concetto di incentivo: in realtà, anche nel contesto neoclassico non viene esclusa a priori la possibilità che vi sia informazione distribuita in modo asimmetrico.

L’ipotesi che viene fatta, però, è inerente “l’incapacità” dei soggetti di sfruttare a proprio vantaggio (ed anche a svantaggio di altri) tali informazioni; viene, in altri termini, ipotizzato che i soggetti trasmettano queste informazioni senza cercare di trarne vantaggio; Kreps parla di “egoismo non opportunista”⁸.

Nella nuova impostazione, invece, si ipotizza (in modo più realistico) che i soggetti cerchino di trarre vantaggio dalle informazioni in proprio possesso e quindi risulta necessaria la costruzione di un sistema di incentivi volto a far trasferire tale informazione. Il processo di trasferimento, però, è costoso e genera inefficienze.

⁸ Kreps D. M., *Corso di microeconomia*, Il Mulino, p. 867. 1993

Cominciamo ora a trattare più diffusamente la questione informativa; esistono due principali tipi di asimmetrie:

1. informazione nascosta (opportunismo pre-contrattuale), che si verifica prima dello scambio; una delle due parti non conosce tutte le caratteristiche dell'altra, ne segue un fenomeno di selezione avversa, in quanto la parte informata cerca di mostrare le proprie caratteristiche migliori di quello che siano realmente, come visto nell'esempio del mercato delle auto usate;
2. azione nascosta, che si verifica dopo lo scambio; una delle due parti non conosce l'azione, l'impegno e la veridicità delle dichiarazioni dell'altra parte, ne segue un fenomeno di azzardo morale.

In questa analisi intendo concentrarmi in particolar modo sul problema dell'azzardo morale nell'ambito del mercato del lavoro; l'obiettivo è studiare vari modelli che consentano di prevenire comportamenti opportunistici e analizzarne le implicazioni macroeconomiche. Torniamo, però all'analisi preliminare del problema.

La tipica situazione di asimmetria informativa è studiata nell'ambito del modello principale-agente (ad esempio datore di lavoro e lavoratore, manager e azionisti, imprese e banche): l'agente dispone di maggiori informazioni rispetto al principale, informazioni che può usare a proprio vantaggio comportandosi in maniera opportunistica, comportamento a cui non si può risalire a partire dai risultati a causa dell'incertezza; i risultati, cioè, derivano sia dal "lavoro" dell'agente, sia dal contributo di una variabile casuale.

In tal contesto si parla di azzardo morale, detto anche opportunismo post-contrattuale, quando l'agente compie un'azione nascosta che influenza, senza che

tale azione fosse prevista dal contratto (e desiderabile dal principale), l'utilità di entrambe le parti⁹.

Molti sono gli esempi che si riscontrano nella realtà, basti pensare al mercato assicurativo, ma anche alla relazione tra azionisti e manager.

L'asimmetria informativa genera l'interdipendenza delle due parti¹⁰, che rende necessario affiancare al meccanismo di domanda e offerta un'altra metodologia, per l'appunto la teoria dei contratti: si ripresenta, quindi, la dicotomia, già analizzata, fra "transazione di mercato" e "transazione relazionale".

Moltissimi sono stati i contributi teorici in questo ambito e ancora oggi la produzione letteraria continua ad essere abbondante: nonostante l'idea di fondo della contrapposizione di due parti con interessi (quasi) contrapposti sia rimasta in tutta la letteratura scientifica, c'è stato un notevole sviluppo nell'approccio metodologico, che, come vedremo, ha portato anche a significativi risultati a livello macroeconomico.

Questa precisazione è opportuna in quanto alcuni concetti e ipotesi, utilizzate nelle prime impostazioni teoriche, sono state abbandonate, o quantomeno affiancate, in favore di assunzioni più articolate che consentissero ai modelli microeconomici di allinearsi con l'evidenza empirica e di giungere a risultati macroeconomici coerenti con la realtà.

Una delle prime impostazioni teoriche nella teoria dei contratti è rappresentata dalla relazione di agenzia: questo filone d'analisi è nato dall'esigenza di studiare la relazione tra il principale e l'agente e formalizzare la dinamica d'interazione al fine di conciliare i contrapposti interessi delle parti.

⁹ L'azzardo morale si può avere anche in condizioni di simmetria informativa, nel momento in cui l'autorità esterna, volta a imporre il corretto adempimento del contratto, non è in grado di verificare determinate variabili. Questo è il problema che verrà affrontato, in seguito, dai contratti impliciti.

¹⁰ L'utilità delle due parti è negativamente correlate: al crescere di una, l'altra decresce.

I contratti di agenzia, almeno con riferimento alla questione centrale da essi affrontati in origine (cioè il *trade-off* tra incentivo e ripartizione del rischio), non hanno avuto “seguito” nell’analisi macroeconomica, come si vedrà invece per altri filoni di ricerca, pertanto, ai fini di questo lavoro, svolgono solo un ruolo introduttivo del problema: pur cambiando, anche radicalmente, approccio, il problema di fondo è sempre lo stesso, ossia la contrapposizione di interessi.

L’idea base, che verrà approfondita nel paragrafo successivo, è quella di definire un contratto il cui obiettivo è massimizzare l’utilità di un soggetto sotto il vincolo che l’altro soggetto sia incentivato a partecipare (vincolo di razionalità o di partecipazione) e fare in modo tale che l’agente, una volta deciso ad accettare il contratto, sia incentivato a scegliere proprio il comportamento (livello di *effort*) desiderato dal principale (vincolo di compatibilità agli incentivi). Il fatto di massimizzare l’utilità di uno o dell’altro soggetto dipenderà, in linea di principio, dal grado di concorrenzialità del mercato; ad esempio nel mercato del credito, viene massimizzata l’utilità del cliente, data l’ipotesi di concorrenzialità del sistema bancario.

Per completare il quadro introduttivo, è opportuno sottolineare che nella prima fase dello sviluppo teorico, a differenza della produzione successiva, si poneva particolare attenzione alla posizione dei soggetti rispetto al rischio, ossia all’incertezza legata al futuro.

Questo problema scomparirà in molte sviluppi teorici successivi (ad esempio nei salari d’efficienza¹¹), dove il rischio non è più una variabile chiave dell’analisi e ne viene semplicemente ipotizzata la neutralità di tutti i soggetti.

Per capire l’impostazione originaria dei contratti d’agenzia, ma anche il motivo per cui il rischio non è più rilevante in molti lavori successivi, è necessario andare a vedere cosa si intende con avversione al rischio.

¹¹ Shapiro C., Stiglitz J.E., *Equilibrium Unemployment as a Worker Discipline Device*, American Economic Review, pp 433-443. 1984

In estrema sintesi, possiamo dire che un soggetto avverso al rischio, di fronte alla scelta di due alternative, una delle quali produrrà con certezza un determinato risultato¹² mentre l'altra darà lo stesso risultato della precedente ma solo in media, sceglierà l'alternativa certa¹³ in quanto questa fornisce un livello maggiore di utilità.

Per un'analisi più dettagliata del problema delle scelte in condizioni di incertezza, si guardi l'approfondimento contenuto in appendice A.1

Ipotizziamo, da qui in avanti, la neutralità al rischio del principale, che potendo stipulare più contratti può diversificare il rischio, e l'avversione al rischio dell'agente.

Questa ultima ipotesi implica che addossando un maggior rischio all'agente, questo chiederà una maggiore remunerazione attesa, o meglio sarà necessaria una remunerazione attesa maggiore al fine di fargli accettare il rischio.

Per quanto riguarda più strettamente la teoria dei contratti, andiamo a vedere quali sono i caratteri distintivi e le linee guida principali nell'ambito dei contratti d'agenzia.

¹² Per risultato si intende esito, realizzazione di un evento.

¹³ E questo avverrà sempre, tranne nel caso in cui il soggetto avverso al rischio venga remunerato adeguatamente per l'incertezza che si pone a carico.

2.1 Il *trade-off* tra incentivo e ripartizione del rischio: dal *first* al *second best*. I contratti d'agenzia.

Nel tipico contratto di agenzia¹⁴ sono presenti due soggetti: il principale e l'agente. Il principale fa ricorso all'agente per svolgere una determinata attività; entrambe le parti possono essere soggetti individuali oppure enti: l'azione dell'agente contribuisce alla produzione della redditività del principale; l'azione, però, genera disutilità per il lavoratore, che quindi è incentivato a impegnarsi il meno possibile, data l'impossibilità del principale di avere pieno controllo sull'operato dell'agente.

Lo scopo della teoria dei contratti, consiste nel formulare un accordo (contratto) che induca entrambe le parti a fare il bene comune, pur massimizzando solo i propri interessi: l'obiettivo è fare in modo tale che ciascuno “mirando” esclusivamente al proprio interesse, sia “costretto” a fare gli interessi anche dell'altra parte.

Il problema centrale è capire l'interazione tra tutte le possibili variabili in gioco: sarebbe facile, ad esempio, definire un contratto “ottimo” in una situazione priva di incertezza con *enforcement* esogeno¹⁵. In una situazione di simmetria informativa (o di asimmetria con certezza), infatti, risulta possibile giungere alla situazione di così detto *first best*, cioè la situazione migliore in assoluto: quando il principale riesce a controllare, in modo perfetto e pieno, la variabile oggetto dell'operato dell'agente, è possibile costruire un contratto basato sull'*effort*.

In questa situazione (di perfetta e simmetrica informazione) per ottenere l'*effort* che produce il risultato desiderato dal principale è sufficiente che l'agente,

¹⁴ Nicita A., Scoppa V., *Economia dei contratti*, Carocci editore, p 29. 2005

¹⁵ La contrapposizione fra *enforcement* esogeno ed endogeno verrà approfondita nei capitoli successivi: per *enforcement* esogeno si intende la possibilità di ricorrere ad una corte per far valere le clausole di un contratto esplicito.

qualora l'azione desiderata dal principale è effettivamente conseguita, venga remunerato con un opportuno salario costante (w) oppure non riceva niente quando, invece, l'azione dell'agente è diversa da quella desiderata dal principale. In queste circostanze occorre solo garantire la partecipazione dell'agente, cioè fare in modo tale che la remunerazione compensi il costo opportunità del lavoratore.

Nel caso di asimmetria senza incertezza, si può ottenere un risultato equivalente a quello precedente, introducendo un contratto che rende l'agente titolare del reddito residuale (situazione di *residual claimant*), al netto di una quota fissa F che rappresenta il compenso (*payoff*) per il principale.

Con questo contratto, l'agente si trova nella situazione di "*residual claimant*" e diventa pienamente responsabile della prestazione complessiva: il prodotto marginale coincide con l'utilità marginale dell'agente che quindi è incentivato a scegliere l'azione ottimale per se stesso, che coincide con quella del principale. Nonostante l'efficienza allocativa raggiunta grazie a questo contratto, occorre, però, sottolineare che la situazione di *residual claimant* non è empiricamente riscontrabile: non esiste al mondo nessuna impresa che "affitterebbe" l'attività imprenditoriale ai propri dipendenti.

Da ora in poi supponiamo che il principale sia neutrale al rischio, mentre l'agente sia avverso.

Se l'azione dell'agente non è osservabile o verificabile e l'incertezza non permette di procedere a ritroso dal risultato per analizzare l'azione, si viene ad avere un *trade-off* tra l'incentivo e la ripartizione ottimale del rischio e il tipo di contratto precedentemente analizzato (situazione di *residual claimant*) comporta notevoli problemi.

Il rischio, infatti, viene distribuito in modo ottimo se è interamente accollato al principale poiché, per ipotesi, neutrale al rischio e quindi capace di attuare

un'operazione di diversificazione contrattuale; in questo caso, però, l'agente non avrebbe incentivo ad impegnarsi e sceglierebbe l'azione, per lui meno onerosa.

Risulta altresì che gli incentivi sono tanto più forti quanto più il guadagno dell'agente è legato al risultato finale, in questo modo, però, il rischio è totalmente a carico dell'agente, che per ipotesi è avverso al rischio: questa, pertanto, è una situazione inefficiente. Si comprende, quindi, che la situazione "ottimale" (sia nell'ottica allocativa che relativa agli incentivi) che si realizza in pratica non può che essere un compromesso, quella che gli economisti chiamano "soluzione di *second – best*".

Definiamo la funzione di utilità del principale come il valore atteso della differenza tra il risultato (che dipende dall'*effort* e dall'incertezza) e il salario pagato (anch'esso funzione dell'*effort* e dell'incertezza, in quanto collegato al risultato) ¹⁶

$$u_p = E\{ y(e,\theta) - w[y(e,\theta)] \} \quad (2.1)$$

la funzione di utilità dell'agente ¹⁷, invece, è data dal valore atteso dell'utilità derivante dal salario al netto della disutilità ¹⁸ causata dall'*effort*,

$$u_a = E\{ u[w[y(e,\theta)]] - c(e) \} \quad (2.2)$$

Il principale agisce per primo ed è in grado di prevedere la reazione dell'agente e quindi massimizzerà i propri profitti attesi data la reazione dell'agente; una volta trovato il livello e^* , ottimale per il principale e per l'agente, occorre costruire uno schema di pagamento volto a conseguirlo, schema che deve rispettare due vincoli:

¹⁶ Per quanto riguarda l'impianto analitico in questo paragrafo si fa riferimento a : Nicita A., Scoppa V., *Economia dei contratti*, Carocci editore. 2005

¹⁷ Crescente e concava ($u_a' > 0$ e $u_a'' < 0$) per via dell'ipotesi di avversione al rischio.

¹⁸ La funzione $c(e)$, crescente e convessa, indica la disutilità derivante dall'*effort*.

1. vincolo di partecipazione, l'utilità attesa dell'agente deve essere almeno pari all'utilità di riserva (\bar{u}), quindi la (2.2) valutata in e^* deve essere maggiore o uguale a \bar{u} ;
2. vincolo di compatibilità agli incentivi, l'utilità dell'agente derivante dalla scelta di e^* deve essere maggiore o uguale a quella derivante da qualsiasi altro livello di e .

Per il principale, il contratto ottimale è quello che massimizza i profitti all'interno della classe dei contratti che soddisfa i due vincoli. Consideriamo alcuni delle principali caratteristiche del contratto di agenzia:

- la retribuzione “incentivante” è funzione crescente dei risultati;
- il vincolo di partecipazione è stringente, cioè è tendenzialmente verificato come uguaglianza;
- si ottiene un'efficienza di *second best*, tendenzialmente inferiore a quella associata al *first-best*;
- se esistono variabili che danno informazioni circa lo sforzo dell'agente, è opportuno includerle nel contratto.

L'efficienza di *second-best* è la migliore possibile date le “imperfezioni” contrattuali; tale risultato deriva da due tipi di inefficienza:

1. la ripartizione del rischio non avviene in modo ottimale (inefficienza allocativa);
2. il livello di *effort* è minore di quello che si ha nel *first-best*, in quanto, tendenzialmente, per tener conto dell'avversione al rischio degli agenti, l'intensità degli incentivi è attenuata (inefficienza produttiva).

L'abbandono di una situazione di *first-best*, però non è inevitabile, infatti, è possibile raggiungere nuovamente tale risultato, anche in condizioni di asimmetria, se sussistono particolari condizioni:

- la relazione tra *effort* e risultato è deterministica, quindi a partire dai risultati è possibile ricostruire l'azione dell'agente;
- l'agente è neutrale al rischio e quindi non è più costoso accollarsi il rischio stesso (situazione di *residual claimant*); il contratto ottimale prevede che tutto il reddito, al netto del pagamento al principale (indicato precedentemente con F), sia assegnato all'agente;
- il principale può inferire l'azione dell'agente; se esistono eventi la cui realizzazione implica uno sforzo nullo da parte dell'agente, allora è possibile “minacciare” l'agente stesso.

Empiricamente, una tipologia di contratti che ha avuto notevole diffusione sono i contratti con remunerazione lineare.¹⁹

In questo tipo di contratti, l'agente riceve un salario fisso s più una quota b (tra 0 e 1) legata al prodotto finale y

$$w(y) = s + by \quad (2.3)$$

dove b , pertanto, indica l'intensità degli incentivi.

I contratti lineari, nonostante la diffusione e la semplicità, non rappresentano, in linea di principio, contratti efficienti. Nonostante questo limite iniziale, è possibile determinare sotto quali condizioni questo tipo di contratto rappresenta la forma contrattuale maggiormente efficiente; questo avviene se

- la struttura è multiperiodale e l'agente decide stadio per stadio una sequenza di azioni che influenza un risultato cumulativo;
- l'agente conosce i risultati fino allo stato corrente e in base alle azioni pregresse può modificare le sue azioni e determinare il risultato finale.

¹⁹ *Ibid.*, p 59

E' opportuno notare anche che il contratto con remunerazione lineare è un contratto “generale” che consente, in base all'intensità dell'incentivo, di descrivere diverse situazioni: con $b=0$ abbiamo un salario fisso (ottimale, dal punto di vista allocativo, ma non incentivante), con $b=1$ abbiamo la situazione di *residual claimant*²⁰ (ottimale dal punto di vista incentivante, ma inefficiente dal punto di vista allocativo), per tutti gli altri valori di b , abbiamo situazioni intermedie; ovviamente al variare di b varia anche l' “esposizione” del lavoratore al rischio.

Supponiamo che il prodotto dell'impresa dipenda dall'*effort*, ossia

$$y = \pi e + \varepsilon \quad (2.4)$$

dove π è la produttività dello sforzo e ε è un *white-noise*, ovvero una componente di disturbo stocastica, con media nulla e varianza σ^2 che indica il livello di incertezza.

L'utilità del lavoratore, data l'incertezza, viene misurata sulla base dell'equivalente certo (U_c)²¹, cioè la differenza tra il valore atteso e il premio per il rischio

$$U_c = s + b\pi e - c(e) - \frac{1}{2}rb^2\sigma^2 \quad (2.5)$$

dove $(s + b\pi e)$ è il salario atteso, $c(e)$ il costo in termini di utilità del lavoro e l'ultimo termine indica il premio per il rischio, con r che indica il grado di avversione al rischio del lavoratore. Ipotizziamo che $c(e) = \gamma e^2/2$.

In base agli incentivi, l'agente massimizza l'equivalente certo rispetto ad e : dalla condizione del primo ordine ($\partial u/\partial e = 0$) si ricava che il beneficio marginale del

²⁰ Dato che il compenso per il principale, nel caso in cui l'agente sia *residual-claimant*, è $F > 0$, è evidente che nella (2.5) deve risultare $s < 0$.

²¹ *Ibid.*, p 61

maggiore impegno deve essere uguale al costo marginale dello sforzo, ossia deve risultare che

$$b\pi = c'(e) = \gamma e \quad (2.6)$$

Si può, pertanto, trovare il livello di e in funzione della produttività dello sforzo moltiplicato per l'intensità degli incentivi e del prodotto e si vede che lo sforzo dell'agente può essere aumentato tramite un aumento dell'intensità degli incentivi

$$e = \frac{b\pi}{\gamma} \quad (2.7)$$

relazione che altro non è che la funzione di reazione del lavoratore agli incentivi.

Per quanto riguarda i profitti attesi totali dell'impresa, abbiamo che

$$E(\Pi) = E(y - w) = E[(1 - b)y] - s = (1 - b)\pi e - s \quad (2.8)$$

da cui il surplus totale è dato da $S = E(\Pi) + U_c$, cioè la somma dei profitti attesi e dell'equivalente certo.

Se lo sforzo è osservabile, i soggetti possono accordarsi direttamente sul livello ottimale di *effort*. Massimizzando S rispetto ad e risulta che l'*effort* di *first - best* è dato dal livello $e^* = \pi/\gamma$, situazione che si ottiene, nel caso di contratto lineare, quando gli incentivi sono massimi, cioè quando $b = 1$.

In condizioni di asimmetria il contrattato non è basato sul livello di *effort*, ma l'agente può essere incentivato legando la remunerazione ai risultati, ma così si pone in capo all'agente un rischio che deve essere considerato nella formulazione del contratto.

Lo strumento incentivante è b ; sostituendo la funzione di reazione dell'agente in S , e calcolando la condizione del primo ordine rispetto a b , si ottiene

$$b^* = \frac{1}{1 + \frac{r\gamma\sigma^2}{\pi^2}} \quad (2.9)$$

che è tanto maggiore (cioè sono preferibili incentivi forti):

- quanto minore è l'incertezza sulla produzione (indicata da σ^2), quanto maggiore cioè risulta l'accuratezza della misurazione delle performance;
- quanto minore è l'avversione al rischio, se infatti r è molto piccolo, conviene introdurre incentivi molto forti; nel caso estremo di neutralità, la situazione ottima sarebbe $b = 1$;
- quanto minore è l'inverso della reattività dei lavoratori agli incentivi (γ/π) che compare al denominatore di b^* , cioè quanto maggiore è la reattività stessa²²;
- quanto maggiore è il prodotto marginale dello sforzo (π).

In conclusione, la formulazione di b^* ribadisce cioè che è stato detto inizialmente: maggiore è l'avversione al rischio e l'incertezza, minore sarà il parametro b , ossia l'intensità degli incentivi, e quindi l'esposizione dell'agente al rischio; si conferma cioè il *trade – off* tra incentivo e distribuzione ottimale del rischio.

Per un'analisi formale del problema principale – agente, alla luce della teoria dei giochi, si veda l'appendice A.2

²² Questo risultato si spiega alla luce del fatto che non ha senso dare forti incentivi a lavoratori che oltre un certo limite oggettivo non possono impegnarsi (ad esempio operai), mentre è sensato un incentivo ad alto potenziale in un contesto, come quello manageriale, dove l'impegno profuso dai lavoratori può essere significativamente diverso e quindi anche il risultato per l'impresa.

2.2 Osservazioni di sintesi sui contratti d'agenzia

Il contratto d'agenzia, o in generale la relazione d'agenzia, definisce uno scambio tra le parti in causa che generalmente sono due, il principale che delega all'agente una certa attività, dietro ricompensa, svolta in modo discrezionale nell'interesse del principale.

In sintesi possiamo individuare quali sono le caratteristiche chiave del contratto d'agenzia:

- gli interessi delle parti devono essere contrapposti, quindi, ciò che genera utilità per una parte deve creare disutilità per l'altra; in particolare il lavoro genera utilità per il principale poiché determina la sua redditività, ma è disutile per l'agente;
- l'interdipendenza tra le parti genera il rischio di comportamenti opportunistici;
- il principale è neutrale al rischio poiché è in grado, stipulando una pluralità di contratti, di diversificare il "portafoglio" in modo da annullare il rischio; l'agente, invece, è avverso al rischio e quindi tanto maggiore è il rischio a suo carico, tanto maggiore sarà la compensazione richiesta;
- esiste asimmetria informativa; l'agente ha a disposizione più informazioni rispetto al principale, inoltre vi è un livello di razionalità limitata a causa dell'incertezza;
- vi è un costo eccessivamente elevato per valutare l'idoneità dei comportamenti dell'agente;
- l'azione dell'agente determina il risultato voluto dal principale, determina cioè la redditività del principale, tuttavia, a causa dell'incertezza, a partire dal risultato non è possibile ripercorrere il processo a ritroso e analizzare il comportamento dell'agente;

- il principale si deve garantire la partecipazione dell'agente e quindi far in modo tale che l'utilità attesa associata alla remunerazione pagata sia maggiore o uguale all'utilità di riserva (vincolo di partecipazione) e tra tutti i livelli di *effort* occorre fare in modo tale che l'agente scelga quello desiderato dal principale (vincolo di compatibilità con gli incentivi).

Il risultato principale del modello d'agenzia consiste nel collegare la remunerazione dell'agente ai risultati ottenuti, frutto del proprio contributo: in questo modo la remunerazione è variabile e contingente ai risultati osservati.

La remunerazione non dipende direttamente dall'*effort* dell'agente, ma dai risultati che egli stesso contribuisce a produrre; ciò che viene prodotto, però, dipende in gran misura anche da variabili esogene che non rientrano nella sfera d'azione delle parti: per l'agente un contratto di questo tipo è rischioso e pertanto deve essere remunerato per il rischio che si pone a carico.

In altre parole, i contratti così formulati, generano una remunerazione aleatoria per l'agente e ne consegue una riduzione dell'utilità dovuta all'avversione al rischio (introdotta per ipotesi), riduzione che deve essere "compensata", in termini "attesi", dal principale.

La situazione ottimale, dal solo punto di vista dell'allocazione del rischio, si otterrebbe qualora tutto il rischio, legato alla prodotto, fosse a carico del principale, ma l'agente non avrebbe incentivo a impegnarsi; questa sarebbe la situazione ottima anche dal punto di vista sociale, ma in questo modo verrebbero eliminati tutti gli incentivi. Viceversa, avremmo una totale inefficienza (allocativa) ponendo tutto il rischio a carico dell'agente, anche se in questo caso, ovviamente, verrebbe garantita la sua totale lealtà (efficienza produttiva).

Nella realtà economica non vi è solo la prospettiva allocativa; anche la scelta dell'*effort* dell'agente rileva dal punto di vista del benessere sociale per cui deve

essere considerata in un'ottica di efficienza "generale" (che non comprenda, cioè, solo il problema allocativo del rischio).

A seguito di questa precisazione, il concetto di efficienza assume una valenza più ampia; non potendo raggiungere contestualmente l'efficienza allocativa e quella legata agli incentivi, la situazione di "*second-best*" di fatto, è la situazione "socialmente efficiente" da entrambi i punti di vista, dovuta sia all'inefficienza allocativa che a quella produttiva.

Il *second-best*, quindi, non dipende solo dal problema della ripartizione del rischio, infatti, in generale, il livello di *effort* è inferiore a quello che si otterrebbe nella situazione di *first-best*.

Come in ogni circostanza in cui il rischio viene trasferito tra le parti, anche in questo caso abbiamo un costo di trasferimento, che è, di fatto, il costo del contratto.

La più significativa caratteristica della relazione d'agenzia, sta forse nel fatto che il principale non sceglie quale remunerazione dare all'agente (come sarà invece in altri schemi di incentivazione): il datore di lavoro, in questo schema, infatti, si "limita" a definire una differente remunerazione per ogni differente livello di prodotto; in sostanza crea un "menù" nel quale indica per ogni livello di prodotto finale quanto è disposto a pagare.

La relazione salario-*effort* è, però, "filtrata" dalla variabile casuale che incide sul prodotto e pertanto, nonostante la profusione di un livello alto di *effort*, il prodotto può essere "basso" e questo fa sì che anche la remunerazione associata lo sia. Alla fin fine, quindi, il tipo di remunerazione più conveniente per l'agente, in questo contesto, è un salario costante che lo "ripari" dal rischio di impegnarsi "inutilmente", ma che potrebbe indurlo a non impegnarsi affatto.

I contratti d'agenzia, con output osservabile e verificabile, rappresentano un'importante primo passo verso l'analisi della relazione tra principale e agente;

questo approccio si basa fortemente sull'idea che sia possibile costruire un contratto basato indirettamente sullo sforzo dei lavoratori, legando la remunerazione al prodotto, “frutto” dell'*effort*.

Va infine sottolineato nuovamente che la formulazione dei contratti d'agenzia non ha avuto seguito all'interno dell'analisi macroeconomica nel mercato del lavoro, quantomeno in relazione alla problematica di fondo riguardante l'allocazione del rischio.

Come vedremo successivamente, l'idea centrale dei contratti d'agenzia, verrà abbandonata: i salari d'efficienza, i contratti *self-enforcing* (fondati sul concetto di reputazione) e i tornei si basano sull'idea che sia possibile per il datore di lavoro formulare un contratto basato direttamente ed esclusivamente sull'*effort*. Verranno, quindi, abbandonati i contratti costruiti sulla base di indicatori delle *performance* osservabili e verificabili; in questo nuovo scenario sarà, inoltre, possibile sviluppare considerazioni più approfondite a livello macroeconomico.

Dopo aver fatto un quadro introduttivo della problematica di fondo, è necessario andare ad analizzare i nuovi filoni teorici, introducendo a tale scopo i salari d'efficienza, i contratti impliciti e successivamente e i tornei.

Appendice

A.1 Scelte in condizioni di incertezza: la posizione dei soggetti nei confronti del rischio (cenni)

Consideriamo un soggetto che, date alcune alternative possibili, deve prendere una decisione relativa ad un evento incerto. Ipotizziamo che il soggetto conosca l'insieme dei possibili m risultati, intesi per ipotesi come somme monetarie, (W_1, \dots, W_m) ed anche la probabilità di ciascun risultato (p_1, \dots, p_m) ; il generico risultato W_k , si realizzerà con probabilità p_k .

A questo punto, data un'azione scelta dal soggetto, possiamo definire il concetto di lotteria L come il vettore casuale

$$L \equiv (W_1, \dots, W_m, p_1, \dots, p_m)$$

che riassume l'insieme dei risultati e le loro rispettive probabilità. Il valore atteso della lotteria sarà dato da

$$E(W) = p_1 W_1 + \dots + p_k W_k \quad (2.10)$$

valore atteso che può essere maggiore o uguale a zero (in questo ultimo caso si parla di lotteria attuariale equa) e che consente di ordinare le lotterie.

Definiti questi concetti preliminari, passiamo a considerare quello di utilità attesa. Sia U la funzione di utilità del soggetto che associa un certo livello di utilità alla generica somma monetaria W ; la funzione di utilità attesa viene definita formalmente in questi termini

$$E(U) = U(L) = p_1 U(W_1) + p_2 U(W_2) + \dots + p_k U(W_k) \quad (2.11)$$

cioè la media delle utilità calcolate nei singoli eventi, ponderate per le rispettive probabilità di questi ultimi. Date due lotterie L' e L'' , il soggetto preferirà strettamente L' a L'' se e solo se vale

$$U(L') > U(L'')$$

Risulta evidente che il “valore” attribuito da un soggetto alle due lotterie, oltre che dal valore attuale dipende anche dalla forma e dall’andamento di U .

Data l’ipotesi semplificatrice che vede W come somme monetarie, ipotizziamo che la funzione di utilità sia sempre crescente rispetto ai risultati, cioè

$$U'(W) = \partial U(W) / \partial W > 0$$

Il passo successivo consiste nello studiare gli atteggiamenti dei soggetti nei confronti del rischio; dobbiamo in primo luogo determinare quale sarebbe la situazione certa che rende il soggetto indifferente tra scegliere tale situazione e partecipare alla lotteria, questa particolare situazione prende il nome di equivalente certo (CE_L)

$$U(CE_L) = E[U(W)] \quad (2.12)$$

Definiamo, inoltre, il concetto di premio per il rischio ξ , ossia quella somma monetaria che il soggetto è disposto a pagare per ottenere con certezza il valore medio della lotteria invece che parteciparvi; detto in altri termini il premio è il “prezzo” che un soggetto è disposto a pagare pur di “eliminare” il rischio a proprio carico.

Data una lotteria L che in media farà ottenere al soggetto una somma pari a $E(W)$, il premio per il rischio ξ deve soddisfare la seguente relazione

$$U[E(W) - \xi] = E[U(W)] \quad (2.13)$$

Dalla (2.12) e dalla (2.13) si ottiene che

$$U [E(W) - \xi] = U (CEL)$$

e quindi invertendo la U (possibile in quanto la funzione è monotona crescente), si ricava che $E[W] - \xi$ è l'equivalente certo.

Risulta intuitivo comprendere che maggiore è ξ , maggiore è l'avversione al rischio in quanto maggiore è la cifra che il soggetto è disposto a pagare pur di non essere più soggetto all'incertezza riguardante i flussi reddituale futuri.

Dato che per ipotesi l'utilità è funzione positiva di W (inteso come somma monetaria), l'eventuale positività di ξ implicherebbe che (disuguaglianza di Jensen)

$$U (E[W]) > E[U (W)] \quad (2.14)$$

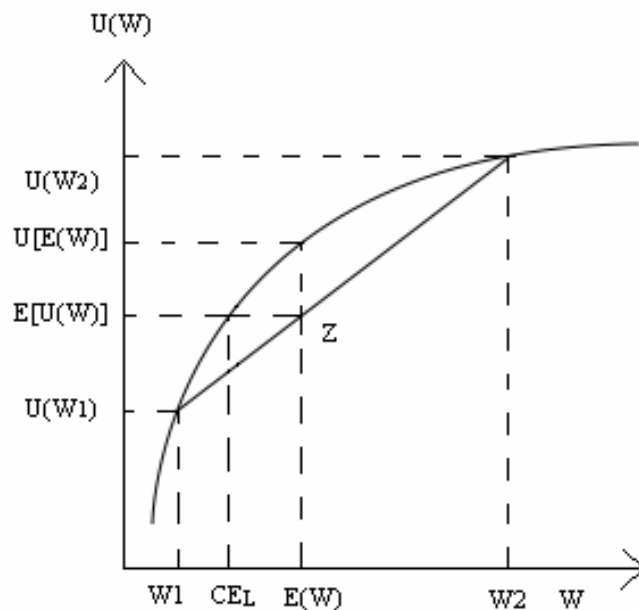
l'utilità di una cifra certa deve essere maggiore rispetto al valore atteso di un reddito incerto che in media ha lo stesso valore della cifra certa.

Quindi se U è concava e la lotteria è una variabile aleatoria con valore atteso finito, vale la disuguaglianza di Jensen. Se, invece, ξ è negativo, cioè un soggetto è disposto a pagare pur di essere sottoposto ad un rischio, allora la funzione di utilità deve essere convessa (propensione per il rischio). Se $\xi=0$ avremo neutralità al rischio.

Per quanto riguarda più in dettaglio la situazione di avversione al rischio, definiamo la seguente lotteria $L = (W_1, W_2; p_1, p_2)$ tale per cui $W_1 < W_2$, il cui valore atteso sarà

$$E(W) = p_1 W_1 + p_2 W_2 \quad (2.15)$$

La funzione di utilità di un soggetto avverso al rischio è sempre crescente rispetto a W ed ha la concavità rivolta verso il basso e da un punto di vista grafico avremo



graf.1

$U[E(W)]$ indica l'utilità del soggetto nel momento in cui ottiene una somma (certa) pari al valore atteso della lotteria. Il segmento che unisce i punti della funzione di utilità in corrispondenza di $U(W1)$ e $U(W2)$ rappresenta, di fatto, la funzione di utilità attesa; tale funzione, infatti, nel caso di due soli possibili risultati, può essere vista graficamente come il segmento precedentemente definito.

Valendo la (2.14), l'utilità che un soggetto avverso al rischio ottiene giocando alla lotteria L , avente come valore atteso $E(W)$, è inferiore all'utilità che il soggetto avrebbe potendo ottenere con una somma (certa) pari al valore atteso della lotteria.

Nel grafico è riportato anche l'equivalente certo (CE_L), cioè la somma monetaria tale per cui vale la (2.12): in questo caso l'equivalente certo è minore del valore atteso della lotteria e quindi, come già detto in precedenza, il premio per il

rischio è positivo. In ultimo possiamo osservare che un soggetto avverso al rischio non accetterà mai una lotteria attuariale equa.

Nel caso di neutralità, la funzione di utilità coincide con la funzione di utilità attesa (sono, cioè, entrambe lineari): il soggetto è indifferente tra ricevere una somma certa oppure giocare alla lotteria. Quindi, il premio è nullo, la funzione di utilità è lineare e l'equivalente certo di una lotteria è uguale al valore atteso e il soggetto è indifferente se accettare o meno lotterie eque.

Nel caso di propensione per il rischio, invece, la funzione di utilità ha la concavità verso l'alto, l'equivalente certo è minore del valore atteso della lotteria e quindi il premio è negativo, inoltre il soggetto accetterà sempre di partecipare a lotterie eque.

Consideriamo un esempio numerico: supponiamo che un investimento possa fruttare 16 con probabilità $\frac{1}{2}$ oppure 0 con probabilità $\frac{1}{2}$. (i risultati dell'investimento vengono intesi come somme monetarie). Ipotizziamo che il soggetto che fa l'investimento sia avverso al rischio e quindi abbia una funzione di utilità tale per cui $U' > 0$ e $U'' < 0$, ad esempio $U(x) = \sqrt{x}$.

Il valore atteso dell'investimento è dato da $\frac{1}{2} * 0 + \frac{1}{2} * 16 = 8$, mentre il valore dell'utilità attesa dell'investimento è dato da $\frac{1}{2} * U(0) + \frac{1}{2} * U(16) = 2$.

Supponiamo di partire da una situazione in cui il capitale è 8, cioè pari al risultato medio dell'investimento: l'utilità che ne deriva è $U(8) = \sqrt{8} > 2$, maggiore cioè dell'utilità attesa dall'investimento. E vero che l'investimento, in media, darà 8, ma è anche vero che l'utilità attesa è inferiore rispetto all'utilità che si ha non investendo (che in questo caso si identifica con l'utilità di un evento certo). In questo esempio è stato ipotizzato che il soggetto possieda già un capitale (pari al risultato medio dell'investimento), ma tale ipotesi serve solo a sottolineare la “certezza” del capitale iniziale; la sostanza non cambierebbe interpretando (come fatto a livello teorico) il capitale iniziale come un altro evento che produrrà con certezza il valore medio dell'investimento.

A.2 L'approccio della teoria dei giochi nella relazione principale-agente: schema semplificato d'analisi

Per poter rappresentare in modo analitico²³ il rapporto principale-agente alla luce della teoria dei giochi, è necessario indicare l'insieme dei giocatori, le strategie e i *payoff*. I giocatori sono due: il lavoratore (l'agente) che per ipotesi è avverso al rischio e il datore di lavoro (il principale) neutrale al rischio. Indichiamo con S l'insieme di scelta del lavoratore

$$S = \{e_1, e_2\} \text{ con } e_2 > e_1 > 0. \quad ^{24}$$

Come precedentemente ipotizzato, il prodotto y , verificabile da parte del principale, dipende dall'*effort* del lavoratore ma anche da una variabile aleatoria β , non verificabile, che supponiamo possa assumere il valore β_1 con probabilità p e β_2 con probabilità $1-p$.

Nell'approccio precedente era stato ipotizzato che la componente stocastica che influisce sul prodotto²⁵, fosse una variabile casuale con media nulla e varianza σ^2 ; ora, invece, si sta operando una forte semplificazione che consente di affrontare il problema sempre da un punto di vista formale, ma con maggior facilità.

In sostanza, si sta ipotizzando che la variabile aleatoria che incide sul prodotto si concretizzi solo tramite due particolari realizzazioni che combinate con un certo livello di *effort* producono un certo livello di prodotto; ovviamente i giocatori non conoscono quale sia il valore realizzato della variabile casuale, ma conoscono la probabilità della loro realizzazione e l'incidenza, combinata allo sforzo, sul prodotto.

Supponiamo che il prodotto assuma questi valori in funzione dell'*effort* e di β

²³ Grillo M., *Informazione asimmetrica e incentivi – il problema del moral hazard*, Note didattiche, Facoltà di scienze politiche di Pisa. 1991.

²⁴ Il livello $e = 0$ viene implicitamente considerato: in questo caso il gioco termina.

²⁵ A causa della quale non è possibile risalire all'*effort* a partire dai risultati

y	β_1	β_2
e1	10	10
e2	10	40

La distribuzione sugli stati del mondo $p(\beta)$ induce ad una distribuzione π sul prodotto y , che dipende dallo sforzo dell'agente, $\pi = \pi(y,e)$, quindi, con $y_1=10$ e $y_2=40$, avremo $\pi(y_1,e_1)=1$, $\pi(y_2,e_1)=0$, $\pi(y_1,e_2)=p$, $\pi(y_2,e_2) = 1-p$.

E' opportuno chiarire questo punto: come ipotizzato nella tabella il prodotto può assumere solo due valori (10, 40) ma in realtà le combinazioni che producono tali valori non sono soltanto due bensì quattro. L'espressione $\pi(y_1,e_1)=1$ sta ad indicare che con il livello di *effort* e_1 , la probabilità che il prodotto assuma valore y_1 è unitaria (cioè è un evento certo) e questo è evidente dalla tabella stessa. Similmente possiamo dire che:

- $\pi(y_2,e_1) = 0$ indica che con e_1 la probabilità di ottenere y_2 è nulla;
- $\pi(y_1,e_2) = p$ indica che con e_2 la probabilità di ottenere y_1 è pari alla probabilità di realizzazione di β_1 , ossia p ;
- $\pi(y_2,e_2) = 1-p$ indica che con e_2 la probabilità di ottenere y_2 è pari alla probabilità di realizzazione di β_2 , ossia $1-p$.

Sintetizzando le considerazioni appena svolte possiamo rappresentare in forma tabellare le probabilità con cui si realizzano i diversi livelli di prodotto in base al livello di *effort*

$\pi(y,e)$	e1	e2
y1	1	p
y2	0	1-p

Al fine di poter usare la variabile prodotto (verificabile) come *proxy* dell'*effort* (non verificabile), occorre ipotizzare che la probabilità che il prodotto y sia non

maggiore di qualsiasi altro livello arbitrario, è tanto più grande quanto più piccolo è il livello d'effort (dominanza stocastica del primo ordine).

Per quanto concerne il principale, il datore di lavoro pagherà all'agente, w_1 se $y=y_1$ e w_2 se $y=y_2$; pertanto l'insieme strategico del principale è dato dalle coppie (w_1, w_2) tali che

$$w(y) = \begin{cases} w(y_1) = w_1 \\ w(y_2) = w_2 \end{cases}$$

Definiti gli insiemi strategici, occorre definire cosa si intende per contratto; un contratto, fondamentalmente, non è altro che un insieme di tre elementi e può essere indicato in questi termini $[(w_1, w_2); e]$, ossia una combinazione tra un meccanismo di pagamento e la scelta dell'agente in termini di effort; il gioco si sviluppa in questo modo: il principale sceglie il metodo di pagamento e l'agente, in base al meccanismo di pagamento, sceglie lo sforzo.

La funzione di utilità dell'agente è $U_a = U_a(w, e)$ tale per cui $\partial U_a / \partial w > 0$ (salario accresce l'utilità), $\partial^2 U_a / \partial w^2 < 0$ (il soggetto è avverso al rischio), $\partial U_a / \partial e < 0$ (l'effort è disutile per l'agente).

Una specificazione funzionale che rispetta le ipotesi relative alla funzione di utilità dell'agente è data da

$$U_a = \sqrt{w} - e \quad (2.16)$$

e la funzione di *payoff* dell'agente è la sua utilità attesa definita rispetto all'insieme di tutti i possibili contratti, ossia

$$EU_a = EU_a [(w_1, w_2); e] \quad \text{tale che}$$

$$EU_a = \begin{cases} \pi(y_1, e_1) \sqrt{w_1} + \pi(y_2, e_1) \sqrt{w_2 - e_1} & \text{se } e = e_1 \\ \pi(y_1, e_2) \sqrt{w_1} + \pi(y_2, e_2) \sqrt{w_2 - e_2} & \text{se } e = e_2 \end{cases} \quad (2.17)$$

da cui

$$EU_a = \begin{cases} \sqrt{w_1 - e_1} & \text{se } e = e_1 \\ p \sqrt{w_1} + (1-p) \sqrt{w_2 - e_2} & \text{se } e = e_2 \end{cases} \quad (2.18)$$

Al fine di semplificare l'analisi numerica indichiamo con $U_a^\circ = 0$ l'utilità di riserva e poniamo $e_1=1$, $e_2=2$.

Per quanto riguarda la funzione di utilità del principale, data la neutralità al rischio, ipotizziamo

$$V_p = y - w(y) \quad (2.19)$$

La funzione di *payoff* è data dalla quella di utilità attesa definita rispetto tutti i possibili contratti,

$$EV_p = EV_p[(w_1, w_2); e] \text{ tale che}$$

$$EV_p = \begin{cases} \pi(y_1, e_1)(e_1 - w_1) + \pi(y_2, e_1)(y_2 - w_2) & \text{se } e = e_1 \\ \pi(y_1, e_2)(y_1 - w_1) + \pi(y_2, e_2)(y_2 - w_2) & \text{se } e = e_2 \end{cases} \quad (2.20)$$

da cui

$$EV_p = \begin{cases} y_1 - w_1 & \text{se } e = e_1 \\ p(y_1 - w_1) + (1-p)(y_2 - w_2) & \text{se } e = e_2 \end{cases} \quad (2.21)$$

e, per semplicità, rappresentiamo l'utilità di riserva del principale con $V_p^\circ = 0$

Il contratto $[(w_1^*, w_2^*); e^*]$ rappresenterà un equilibrio di Nash se soddisfa le condizioni

$$EV_p[(w_1^*, w_2^*); e^*] \geq EV_p[(w_1, w_2); e^*] \quad (2.22)$$

per ogni $(w_1, w_2) \neq (w_1^*, w_2^*)$

sotto i vincoli, rispettivamente, di compatibilità agli incentivi e di partecipazione

$$EU_a[(w_1^*, w_2^*); e^*] \geq EU_a[(w_1^*, w_2^*); e] \quad \text{per ogni } e \neq e^* \quad (2.23)$$

$$EU_a[(w_1^*, w_2^*); e^*] \geq U_a^o = 0 \quad (2.24)$$

Nell'esempio, un contratto è efficiente se e solo se la remunerazione è indipendente dal prodotto, in altri termini se e solo se $w_1 = w_2$. In questo modo, però, l'agente profonderà il livello minimo di sforzo.

Se l'agente, infatti, sceglie e_1 , allora $y = y_1 = 10$ e il principale ha convenienza a proporre un contratto $w_1 = w_2 = w$ tale per cui $\sqrt{w^o} - e_1 = 0$, cioè $w^o = e_1^2$ da cui $w^o = 1$ e i *payoff* saranno rispettivamente $V_p = 10 - 1 = 9$, $U_a = 0$.

Ovviamente la situazione precedentemente descritta non è pareto-efficiente, è possibile, infatti, determinare un miglioramento per entrambe le parti: supponiamo che il principale proponga all'agente una compartecipazione ai profitti del 50% e che questo induca l'agente a profondere $e_2 = 2$.

Il prodotto sarà pari a $y_1 = 10$ con probabilità p e a $y_2 = 40$ con probabilità $1 - p$. Il principale pagherà all'agente $w_1 = 5$ (trattenendo per se un *payoff* pari a 5) con probabilità p e $w_2 = 20$ (trattenendo per se un *payoff* pari a 20) con probabilità $(1 - p)$, da cui i valori attesi dei *payoff*

$$EV_p = 5p + 20(1 - p) = 20 - 15p$$

$$EU_a = \sqrt{5} p + \sqrt{20}(1 - p) - 2$$

Per $p < 11/15$, quindi, questo contratto è superiore, in termini paretiani, a quello che comporta $V_p=9$ e $U_a=0$, ma per quanto superiore al contratto $[(1,1);1]$, è sempre inefficiente in termini paretiani.

Tuttavia, come in parte visto nel paragrafo precedente, una situazione pareto-efficiente resta possibile

1. se l'effort è verificabile: in tal caso il principale paga un certo salario se $e = e_2$ e paga un salario inferiore se $e \neq e_2$, e il salario costante è pagato indipendentemente dalla realizzazione della variabile aleatoria, se $e = e_2$, per cui l'allocazione del rischio è efficiente;
2. se esistono valori della variabile verificabile che si realizzano con probabilità strettamente positiva solo se lo sforzo dell'agente è quello desiderato dal principale;
3. se l'agente è neutrale al rischio.

Per quanto riguarda più in dettaglio quest'ultimo punto, la funzione di utilità dell'agente può essere rappresentata, nel caso di neutralità al rischio, in questo modo $U_a = w - e$, da cui i *payoff* dell'agente e del principale sono rispettivamente dati da

$$EU_a = \begin{cases} w_1 - e_1 & \text{se } e=e_1 \\ pw_1 + (1-p)w_2 - e_2 & \text{se } e=e_2 \end{cases} \quad (2.25)$$

$$EV_p = \begin{cases} y_1 - w_1 & \text{se } e=e_1 \\ py_1 + (1-p)y_2 - pw_1 - (1-p)w_2 & \text{se } e=e_2 \end{cases} \quad (2.26)$$

Supponiamo che il principale possa indurre l'agente a approfondire il livello di effort e_2 , il vincoli di partecipazione dell'agente sarà

$$pw_1 + (1 - p)w_2 - e_2 = U_a^0 = 0 \text{ da cui}$$

$$pw_1 + (1 - p)w_2 = e_2 \quad (2.27)$$

e sostituendola nella funzione di payoff del principale

$$EV_p = py_1 + (1 - p)y_2 - e_2 \quad (2.28)$$

Supponiamo che l'agente diventi titolare del reddito residuale e che paghi tale diritto k , pari al suo *payoff* atteso, ossia

$$k = py_1 + (1 - p)y_2 - e_2 \quad (2.29)$$

Il beneficio netto atteso dell'agente, pagando k , è

$$EB_a = \begin{cases} y_1 - k - e_1 & \text{se } e=e_1 \\ py_1 + (1 - p)y_2 - k - e_2 & \text{se } e=e_2 \end{cases} \quad (2.30)$$

sostituendo il valore di k ,

$$EB_a = \begin{cases} (1 - p)(y_1 - y_2) + e_2 - e_1 & \text{se } e=e_1 \\ 0 & \text{se } e=e_2 \end{cases} \quad (2.31)$$

se p (la probabilità che il prodotto y sia uguale a 10 nonostante il livello di effort profuso dall'agente sia e_2) non è troppo grande, allora avremo $EB_a(e_2=2) > EB_a(e_1=1)$ e l'agente è indotto a scegliere lo sforzo s_2 (giustificazione *ex-post* dell'ipotesi fatta).

Viste le condizioni sotto le quali è possibile ottenere una situazione di *first – best*, bisogna “rassegnarsi” ed accettare il fatto che il più delle volte l'unica soluzione

possibile è un “ottimo di *second-best*”, in quanto l’allocazione del rischio avviene per lo più in modo inefficiente.

Ricaviamo la soluzione di *second-best*: tornando all’ipotesi di avversione al rischio del lavoratore, il principale può formulare un contratto che induca l’agente a scegliere e_2 nel modo seguente

$$\max_{(w_1, w_2)} EV_p(e=e_2) = p(y_1 - w_1) + (1 - p)(y_2 - w_2) \quad (2.32)$$

sotto i vincoli, rispettivamente di compatibilità agli incentivi (IC) e di partecipazione, o razionalità (IR)

$$\begin{aligned} \text{IC} \quad & p\sqrt{w_1} + (1 - p)\sqrt{w_2} - e_2 \geq \sqrt{w_1} - e_1 \\ \text{IR} \quad & p\sqrt{w_1} + (1 - p)\sqrt{w_2} - e_2 \geq 0 \end{aligned} \quad (2.33)$$

Il vincolo di compatibilità agli incentivi, in sostanza, sta dicendo che l’utilità dell’agente, calcolata nel livello di *effort* desiderato dal principale e data la probabilità che nonostante la profusione di e_2 il salario percepito sia w_1 , deve essere maggiore o uguale a quella che il lavoratore otterrebbe impegnandosi e_1 .

Il vincolo di partecipazione è facilmente interpretabile: l’utilità “attesa” al netto di e_2 deve necessariamente essere maggiore o uguale all’utilità di riserva che per ipotesi è nulla. Dall’insieme dei vincoli (2.33) si ricava algebricamente

$$\begin{aligned} \text{IC}' \quad & (1 - p)(\sqrt{w_2} - \sqrt{w_1}) \geq e_2 - e_1 \\ \text{IR}' \quad & p\sqrt{w_1} + (1 - p)\sqrt{w_2} \geq e_2 \end{aligned} \quad (2.34)$$

Per ipotesi $e_2 > e_1$, quindi IC' è soddisfatta solo se $w_2 > w_1$. Occorre sottolineare inoltre che entrambi i vincoli devono essere stringenti, altrimenti sarebbero possibili incrementi paretiani in particolare nel caso di un calo proporzionale di w_1 e w_2 .

Il problema, quindi, consiste nel massimizzare la (2.32) sotto i vincoli (2.34) soddisfatti come uguaglianza.

A questo punto la procedura standard, per trovare il punto di ottimo, prevede la massimizzazione vincolata, ma nell'esempio la procedura è particolarmente semplice grazie ai valori numerici assegnati ai due diversi livelli di *effort*, da cui otteniamo

$$\begin{aligned} IC'' & (1 - p) (\sqrt{w_2} - \sqrt{w_1}) = 1 \\ IR'' & p\sqrt{w_1} + (1 - p) \sqrt{w_2} = 2 \end{aligned}$$

e sottraendo IC'' da IR'' avremo $p\sqrt{w_1} + (1 - p)\sqrt{w_1} = 1$ da cui $\sqrt{w_1}=1$ e quindi risulterà

$$\begin{aligned} w_1 &= 1 \\ w_2 &= [(2 - p)/(1 - p)]^2 \end{aligned}$$

soluzione che risulta evidentemente non di *first-best* in quanto, come precedentemente sottolineato, data l'avversione al rischio dell'agente, il contratto ottimo in assoluto prevede $w_1 = w_2$.

3. Salari d'efficienza: le origini della teoria, profilo generale e modello di Shapiro e Stiglitz. Profili salariali legati all'anzianità (cenni)

L'impostazione dei salari d'efficienza è stata sviluppata nell'ambito della nuova macroeconomia keynesiana (NMK): questa scuola è nata dalla contrapposizione rispetto alle precedenti impostazioni teoriche ed in particolare rispetto alla nuova macroeconomia classica (NMC) che aveva lanciato un'importante sfida, cioè creare un modello macroeconomico che partisse da solide basi microeconomiche.

La nuova macroeconomia keynesiana è nata, tra le altre cose, dall'esigenza di fornire una spiegazione razionale e microfondata alla rigidità salariale, registrata nel mercato del lavoro il quale veniva visto, a differenza di alcune impostazioni precedenti, come un mercato con caratteristiche peculiari.

In altri termini, la NMK iniziò ad analizzare il mercato del lavoro in sé, abbandonando l'idea neoclassica che il lavoro fosse assimilabile ad una merce; questa scuola, inoltre, si distanziò dal paradigma secondo il quale il livello d'occupazione è (soltanto) “frutto” delle *performance* del sistema economico.

La NMK (ri)pose con forza al centro del dibattito teorico la disoccupazione involontaria e le fluttuazioni macroeconomiche. L'aspetto che più interessa ai fini dell'analisi qui svolta è, in particolare, il contributo della NMK nella spiegazione, basata sulla rigidità del salario, della persistenza della disoccupazione involontaria.

Gli economisti della NMK ritenevano che l'aggiustamento dei salari non fosse così rapido da consentire in ogni istante l'equilibrio all'interno del mercato del lavoro; la vischiosità salariale non consente quindi di riassorbire la

disoccupazione: il nucleo centrale dell'analisi della NMK consiste nel giustificare il fatto che il salario reale, nella pratica, non scende (nonostante la pressione competitiva dei lavoratori) e che la disoccupazione involontaria permane.

La teoria dei salari d'efficienza – o per meglio dire le teorie dei salari d'efficienza rappresentano uno dei più importanti punti di riferimento all'interno della NMK per quanto riguarda la microfondazione delle rigidità salariali.

Come già detto, la NMK pone nuovamente al centro del dibattito teorico la disoccupazione; i salari d'efficienza rappresentano, pertanto, una teoria della disoccupazione che, però, si distanzia dalle altre in quanto l'oggetto d'analisi è la “disoccupazione involontaria d'equilibrio”¹, espressione che può risuonare contraddittoria ma che in realtà non lo è.

L'espressione “disoccupazione involontaria” non necessita di essere ulteriormente chiarita; per quanto riguarda, invece, il termine “equilibrio” è opportuno chiarirne il significato nel contesto della frase.

Come vedremo tra poco, la teoria dei salari d'efficienza, consiste nel trovare un modo per evitare lo *shirking* (cioè l'inattività) dei lavoratori e, come vedremo, questo “modo” sarà il risultato di azioni perfettamente razionali da parte dei singoli individui, razionali, cioè, dal punto di vista microeconomico.

In sintesi possiamo dire che la “razionalità profusa” nelle scelte microeconomiche comporta rendite per i lavoratori occupati², ma allora potremo chiederci il motivo per cui le imprese non assumono *outsider* ad un prezzo inferiore giungendo quindi al pieno impiego.

¹ Brucchi L., *op. cit.*, p 313

² Essere licenziati, quindi, implica la perdita di tali rendite: il licenziamento è costoso ed è reso ancor più costoso dalla presenza di disoccupazione involontaria dovuta proprio alla presenza di rendite per i lavoratori occupati.

In altre parole: perché a comportamenti razionali microeconomici sembrerebbe non corrispondere una razionalità collettiva³?

Come vedremo, non è semplice rispondere a questa domanda, quantomeno se si cerca di farlo da un'ottica esterna al modello; possiamo tuttavia dire che il salario d'efficienza gioca un duplice ruolo, quello incentivante all'interno dell'impresa e quello allocativo all'interno del mercato.

La prima di queste funzioni, però, è incompatibile con il ruolo di variabile equilibratrice del mercato del lavoro; è questa incompatibilità che genera la disoccupazione involontaria⁴.

A questo punto risulta necessario porsi anche un'altra importante domanda: i vari modelli introdotti, finalizzati a microfondare la rigidità salariale sono indubbiamente significativi e forniscono una spiegazione del tutto razionale e valida, ma di per sé sono sufficienti a dimostrare che la rigidità salariale è fonte di disoccupazione?

Gli economisti della nuova macroeconomia keynesiana, infatti, affermando che una causa della disoccupazione è la rigidità salariale, implicitamente stanno asserendo che la flessibilità salariale potrebbe consentire di giungere e permanere alla piena occupazione, anche se in realtà lo stesso Keynes non riteneva la flessibilità uno strumento adatto per il raggiungimento di tale situazione⁵.

Per concludere il quadro introduttivo relativo alla NMK, all'interno della quale prese vita sia la teoria dei contratti impliciti che quella dei salari d'efficienza, è opportuno sottolineare nuovamente che questa scuola cercò di analizzare il mercato del lavoro tenendo conto, a differenza di scuole precedenti,

³ *Ibid.*, p 313

⁴ *Ibid.*, p 314

⁵ A tal proposito si rimanda alle spiegazioni contenute nel primo capitolo della tesi.

dell'esistenza di incertezza e asimmetrie informative fornendo un paradigma scientifico più realistico, ma non per questo esente da critiche.

Inoltre, va osservato preliminarmente che, a differenza dei contratti d'agenzia, in assenza di misure oggettive delle prestazioni e in presenza di osservabilità imperfetta e non verificabilità dell'*effort*, risulta impossibile far ricorso ad una corte per far valere le clausole contrattuali. Come vedremo, il lavoratore sarà leale per via della minaccia di disoccupazione; l'impresa, invece, non avrà motivi per essere sleale dal momento che il salario d'efficienza, una volta stabilito, può essere facilmente definito in modo esplicito oppure può essere pagato in anticipo⁶.

Prima di procedere occorre sottolineare che la teoria dei salari d'efficienza può essere ricondotta a due principali impostazioni: la prima è rappresentata dal modello di Solow, la quale fornisce soltanto un quadro introduttivo delineando alcuni tra i principali aspetti caratterizzanti della nuova impostazione. La "seconda impostazione" si può identificare con il modello di Shapiro–Stiglitz, il quale riprende le idee del modello di Solow approfondendole ed estendendole.

Passiamo ora a definire più dettagliatamente cosa si intende per salari d'efficienza e ad analizzare i principali contributi teorici.

⁶ Nicita A., Scoppa V., *op. cit.*, p 125

3.1 Salari d'efficienza: profilo generale, modello di Solow e di Shapiro – Stiglitz

La disoccupazione può essere studiata da molte prospettive; un primo approccio consiste nello studio delle cause e degli effetti, in modo da valutare se la disoccupazione rappresenta un fallimento del mercato. Le cause, in particolare, possono essere molteplici: possiamo vedere la disoccupazione come strutturale o frizionale; da un'altra prospettiva, invece, possiamo vederla come risultato delle caratteristiche *non-walrasiane* del sistema economico

Un'altro modo per studiare la disoccupazione consiste nell'analizzarne il comportamento ciclico: il salario reale è aciclico, o al più moderatamente prociclico; un equilibrio walrasiano può essere possibile, ad esempio, se l'offerta di lavoro è molto elastica, ma questa ipotesi non è empiricamente sostenibile e pertanto si arriva ad una situazione, oggetto dell'analisi in questo lavoro, in cui l'equilibrio di mercato non è walrasiano⁷.

La teoria dei salari d'efficienza, si inserisce all'interno del quadro d'analisi della disoccupazione, volgendosi, in particolare, al filone avente come obiettivo l'individuazione delle cause della disoccupazione stessa; i salari d'efficienza, in particolare, sono volti a microfondazione le rigidità salariali.

Uno dei punti chiave dei salari d'efficienza consiste nel fatto che, per l'impresa, pagare un elevato salario rappresenta sia un beneficio che un costo; in altri termini, si viene a configurare un *trade-off*: pagare un salario elevato, infatti, comporta un aumento dei costi di produzione ma anche un aumento delle capacità di sostentamento del lavoratore che quindi lavorerà "meglio"⁸.

⁷ Romer D., *Advanced Macroeconomic*, MacGraw-Hill, p 437. 2006

⁸ *Ibid.*, p 439

Nell'ottica neoclassica, invece, una politica di alti salari genera solo un aumento dei costi; in realtà un opportuno aumento salariale non è detto che riduca i profitti, in quanto può incentivare il lavoratore a impegnarsi, e quindi a produrre, di più. I salari di efficienza sono meccanismi di incentivazione che le imprese usano in mancanza di indicatori oggettivi delle performance e di conseguenza non è possibile il ricorso a un'autorità esterna per assicurare l'*enforcement* contrattuale⁹: l'ipotesi di base è che i lavoratori o si impegnano o fanno *shirking* (oziano).

3.1.1 Schema base nella teoria dei salari d'efficienza: il modello di Solow

In questo modello, viene ipotizzata l'esistenza di un numero elevato (N) di imprese identiche; la funzione di profitti dell'impresa rappresentativa è data da

$$\Pi = pQ - wL \quad (3.1)$$

cioè dai ricavi al netto del monte salario. Per semplificare l'analisi, ipotizziamo che l'unico fattore di produzione sia il lavoro (L) e normalizziamo il prezzo all'unità ($p=1$). Il livello di prodotto Q, di ciascuna impresa, dipende dal numero di lavoratori impiegati, più precisamente la relazione è

$$Q = F(eL), \quad \text{con } F' > 0 \text{ e } F'' < 0 \quad (3.2)$$

dove e rappresenta un parametro d'efficienza ed L le unità di lavoro; eL rappresenta, quindi, le unità di lavoro in termini d'efficienza.

Per ipotesi, il parametro d'efficienza è funzione del salario percepito dai lavoratori e si assume, inoltre, che ne sia funzione positiva: se le imprese

⁹ Nicita A., Scoppa V., *op. cit.*, p 123

aumentano il salario, da un lato ci sarà l'aumento dei costi di produzione, ma si avrà anche un aumento del prodotto dovuto alla maggior efficienza che caratterizzerà le prestazioni dei lavoratori.

Analiticamente¹⁰ il problema viene impostato in questi termini

$$\max_{(L,w)} \Pi = \max_{(L,w)} F[e(w)L] - wL \quad (3.3)$$

da cui, calcolando la condizione del primo ordine rispetto a L e rispetto a W, si ottiene

$$F'[e(w)L]Le'(w) - L = 0 \quad (3.4)$$

$$F'[e(w)L]e(w) - W = 0 \quad (3.5)$$

Dalla (3.5) è possibile ottenere la condizione marginalistica di uguaglianza tra la produttività marginale del lavoro per unità d'efficienza e il costo del lavoro per unità d'efficienza

$$F'[e(w)L] = \frac{w}{e(w)} \quad (3.6)$$

risolvendo il sistema di equazioni (3.4) e (3.6) si ottiene la seguente condizione

$$\frac{w}{e(w)} e'(w) = 1 \quad (3.7)$$

da cui, grazie alla condizione marginalistica di uguaglianza tra la produttività marginale in termini d'efficienza ed il costo del lavoro per unità d'efficienza, possiamo ottenere l'occupazione ottimale: invertendo la (3.6), infatti, si ottiene $L^* = F'^{-1}(w^*/e(w^*))/e(w^*)$, dove w^* è il salario che soddisfa la (3.7).

¹⁰ Romer D., *op. cit.*, p 441

Per comprendere gli effetti del salario d'efficienza nel mercato del lavoro, supponiamo che il livello di salario ottimo sia maggiore rispetto a quello di *market-clealing*; in questa situazione si viene ad avere un eccesso di offerta di lavoro, quindi disoccupazione, ma alle imprese non conviene abbassare il salario altrimenti non massimizzerebbero i profitti, pertanto permane disoccupazione involontaria.

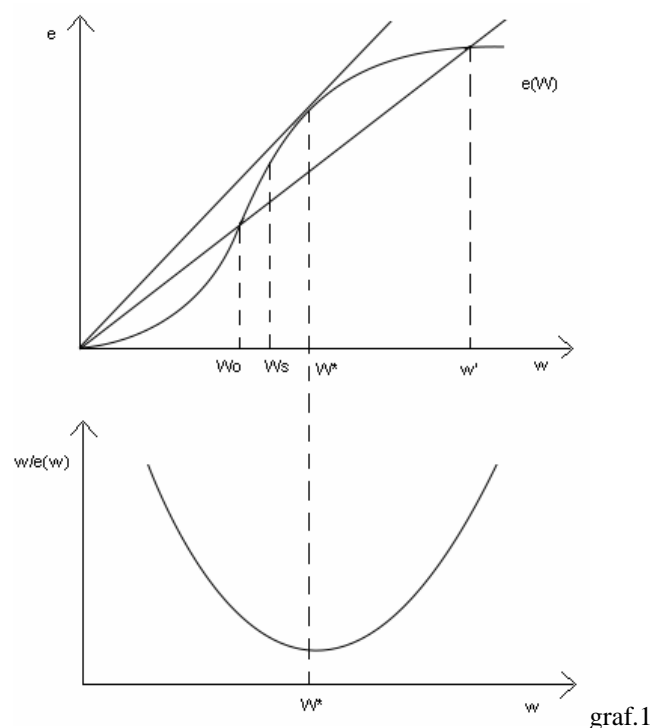
Dal un punto di vista grafico¹¹ è possibile giungere agli stessi risultati ottenuti analiticamente; per quanto riguarda la funzione $e(w)$ supponiamo che

$$e'' > 0 \quad \text{per } w < w_s$$

$$e'' < 0 \quad \text{per } w > w_s$$

questo vuol dire che un aumento del salario, per w minore di w_s , ha “rendimenti” positivi, mentre superato w_s , si vengono ad avere “rendimenti” negativi¹².

Il salario ottimo w^* è quello che minimizza il costo del lavoro per unità d'efficienza $w/e(w)$, da un punto di vista grafico avremo



¹¹ *Ibid.*, p 443

¹² Fino a w_s un aumento del salario si traduce in un aumento più che proporzionale dell'efficienza; superato w_s , invece, l'aumento dell'efficienza diviene meno che proporzionale rispetto all'aumento di w .

consideriamo, infatti, la retta che, a partire dall'origine, interseca la funzione $e(w)$: l'angolo formato dalla retta e dall'asse orizzontale, misura graficamente l'inverso del costo del lavoro per unità di efficienza¹³.

L'angolo aumenta (quindi il costo per unità d'efficienza diminuisce) fino al punto di tangenza, per poi diminuire (quindi il costo per unità d'efficienza aumenta): il massimo (il minimo $w/w(e)$) si ha nel punto di tangenza dove vale

$$\frac{e(w^*)}{w^*} = e'(w^*) \quad (3.8)$$

da cui si ottiene nuovamente la (3.7), per cui l'elasticità della funzione dell'*effort* rispetto al salario è uguale ad uno (condizione di Solow).

3.1.2 Microfondazione della relazione tra salario ed efficienza: l'*effort* come misura dell'efficienza. Modello uniperiodale.

Il modello di Solow rappresenta lo schema base di riferimento all'interno dei contratti d'efficienza, tuttavia è soltanto un approccio introduttivo; la relazione tra salario ed efficienza, infatti, non viene microfondata, ma “semplicemente” ipotizzata. Esistono diversi modi per dare un fondamento a tale relazione, uno di questi consiste nell'ipotizzare che una maggiore efficienza è raggiungibile attraverso un maggior *effort* (impegno lavorativo): questa impostazione è stata utilizzata, in particolare, nel modello di Shapiro e Stiglitz, di cui vediamo una formulazione uniperiodale per poi analizzare successivamente il modello originale intertemporale.

¹³ Da un punto di vista geometrico, la pendenza della retta così costruita è data dal rapporto tra la variabile sull'asse verticale e quella sull'asse orizzontale in corrispondenza del punto in cui le curve si intersecano.

In questa ottica, i salari di efficienza mirano a trovare un livello di remunerazione che, congiunto con il rischio di licenziamento, incentivi ad impegnarsi. In una situazione di equilibrio nel mercato del lavoro, inteso in termini neoclassici (dove cioè il livello salariale è uguale per tutte le imprese e uguale alla produttività marginale e alla disutilità marginale del lavoro), anche se il lavoratore viene licenziato, potrà trovare facilmente un nuovo impiego allo stesso salario.

Se un'impresa paga un salario maggiore di quello d'equilibrio, allora il costo del licenziamento diventa maggiore; se tutte le imprese pagassero un salario maggiore dell'equilibrio, l'incentivo a impegnarsi sembrerebbe scomparire, ma in questo caso avremmo disoccupazione involontaria che rende nuovamente costoso l'eventuale licenziamento¹⁴.

Nei contratti di efficienza, essendo la remunerazione fissa e non legata alle *performance*, né il principale né l'agente hanno incentivo a comportarsi in modo disonesto: il salario di efficienza rappresenta un tipo di contratto con *enforcing* bilaterale¹⁵.

Supponiamo che l'osservabilità sia imperfetta e che un agente possa impegnarsi con *effort* > 0 o nullo e che la sua funzione di utilità sia semplicemente la differenza tra il salario e l'*effort*¹⁶, cioè

$$u = W - e \quad (3.9)$$

definiamo, inoltre, p , appartenente all'intervallo $(0,1]$ ¹⁷, la probabilità di venire scoperti a fare *shirking*.

¹⁴ Nicita A., Scoppa V., *op. cit.*, p 124

¹⁵ *Ibid.*, p 125

¹⁶ La funzione di utilità del lavoratore è lineare, viene, cioè, ipotizzata la neutralità al rischio del lavoratore.

¹⁷ Il fatto che la probabilità p non possa assumere, per ipotesi, il valore zero, sta ad indicare che nella realtà un sistema di controllo totalmente inefficace, non esiste. In questo caso virtuale, il salario d'efficienza avrebbe un valore infinito, per cui non avrebbe senso intraprendere l'attività produttiva. L'ipotesi che p assuma valore unitario è anch'essa irrealistica, ma sicuramente è una situazione a cui è possibile tendere in modo relativamente facile.

In caso di impegno il lavoratore otterrà W^* ; in caso di licenziamento, poiché sorpreso a fare *shirking*, il lavoratore potrà godere al massimo del salario di riserva (W_u). In base all'*effort*, quindi, avremo

$$u = \begin{cases} W^* - e & e > 0 \\ (1 - p)W^* + pW_u & e = 0 \end{cases} \quad (3.10)$$

il lavoratore, quindi, sarà leale se vale $W^* - e \geq (1 - p)W^* + pW_u$; l'impresa, per massimizzare i profitti, pagherà il salario minimo che soddisfi tale disuguaglianza che, quindi, può essere riscritta come uguaglianza stretta

$$W^* - e = (1 - p)W^* + pW_u \quad (3.11)$$

da cui si ricava che il valore del salario d'efficienza è pari¹⁸ a

$$W^* = W_u + \frac{e}{p} \quad (3.12)$$

Si noti che il salario d'efficienza è maggiore del costo opportunità e tanto più basso quanto maggiore è il grado di monitoraggio e controllo p .

Va infine sottolineato che l'evidenza empirica ha fornito alcuni risultati consistenti con la teoria dei contratti d'efficienza, ecco alcuni esempi¹⁹:

- nel 1914 Henry Ford pagò ai suoi dipendenti il doppio del salario di mercato, garantendosi in questo modo un considerevole aumento in termini di produttività;

¹⁸ Nicita A., Scoppa V., *op. cit.*, p 127

¹⁹ *Ibid.*, p 130

- a causa di accordi sindacali un'impresa, con una molteplicità di impianti, fu costretta a pagare lo stesso salario nonostante le diverse realtà economiche locali. I licenziamenti dovuti a comportamenti opportunistici furono tanto minori quanto maggiore era la differenza tra il salario pagato e il salario di riserva nelle diverse zone e quanto più alto era il tasso di disoccupazione locale (Cappelli e Chauvin, 1991);
- Groshen e Krueger (1990) hanno verificato empiricamente che là dove il monitoraggio è minore, i salari sono maggiori: nei *fast-food* il monitoraggio è minimo e salari alti, mentre in un ristorante in *franchising* gestito dal proprietario, il controllo è maggiore e quindi i salari minori;
- è stato riscontrato che alcuni settori, a parità di produttività, pagano salari maggiori a lavoratori con le stesse caratteristiche di quelli in altri settori: questo si ha là dove i costi di monitoraggio sono alti e/o dove il costo per l'impresa di comportamenti opportunistici è elevato.

Un altro studio, di carattere econometrico (Campbell, 1993), basato su un modello di rotazione della manodopera, ha mostrato che, effettivamente, un salario alto riduce la rotazione della manodopera, cosa che, però, non avviene grazie al tasso di disoccupazione²⁰.

3.1.3 Il modello di Shapiro e Stiglitz

Il modello di Shapiro e Stiglitz²¹ è un modello d'azzardo morale basato sull'idea che, data la difficoltà nel controllare l'impegno del lavoratore, i salari d'efficienza possano far giungere all'impegno massimo da parte dei lavoratori, se congiunti con la minaccia di licenziamento.

²⁰ Brucchi L., *op. cit.*, p 326

²¹ Shapiro C., Stiglitz J.E., *Equilibrium Unemployment as a Worker Discipline Device*, American Economic Review, pp 433-443. 1984

Il modello è intertemporale continuo e viene ipotizzato che ciascun operatore abbia a disposizione un arco di tempo *infinito*, e che ciascun lavoratore possa trovarsi, in ogni istante, ad essere occupato o disoccupato.

La funzione di utilità del lavoratore, data l'ipotesi di neutralità al rischio, viene ipotizzata additiva e separabile, nel caso in cui il lavoratore è occupato l'utilità istantanea²² sarà

$$U[W(t), e(t)] = W(t) - e(t) \quad (3.13)$$

dove W è il salario, e indica l'*effort*.

Nel caso in cui il lavoratore è disoccupato, invece, avremo che l'utilità istantanea è rappresentata dall'utilità di riserva che indichiamo con \underline{u} .

A questo punto è necessario calcolare il valore scontato atteso dell'intero arco della vita lavorativa in funzione delle diverse condizioni in cui il lavoratore si trova. Assumiamo che la scelta, e , di impegno sia discreta, e che i casi possibili siano:

$$e = \begin{cases} e > 0 \\ 0 \end{cases}$$

pertanto il lavoratore o si impegna con *effort* positivo oppure ozia.

Indichiamo con V_u il valore scontato atteso della vita lavorativa legato allo stato di disoccupato, V_e^{ns} il valore scontato legato allo stato di occupato che si impegna, V_e^s il valore scontato legato allo stato di occupato che non si impegna²³.

Ai fini dell'analisi, per poter cioè confrontare i valori scontati, sfruttiamo le *asset value equations*²⁴: lo stato del lavoratore viene considerato come un titolo

²² Si parla di utilità istantanea in quanto inseriti in un contesto temporale continuo.

²³ Romer D., *op. cit.*, p 452

²⁴ *Ibid.*, p 451

contrattato da agenti neutrali al rischio in un mercato perfettamente concorrenziale. In equilibrio, il prodotto tra il valore nominale e il tasso di rendimento, sarà uguale al dividendo sommato al guadagno atteso (o a cui vanno sottratte eventuali perdite attese) in conto capitale. Avremo pertanto:

$$r V_u = \underline{u} + a(V_e - V_u) \quad (3.14)$$

$$r V_e^{ns} = W - e - b (V_e^{ns} - V_u) \quad (3.15)$$

$$r V_e^s = W - (b+q) (V_e^s - V_u) \quad (3.16)$$

la (3.14) si riferisce allo stato di disoccupato; r è il tasso di interesse dell'economia, il dividendo è dato dall'utilità istantanea relativa allo stato di disoccupato, cioè dall'utilità di riserva, mentre $a(V_e - V_u)$ rappresenta il guadagno in conto capitale moltiplicato per la probabilità a di trovare lavoro.

La (3.15) riguarda il caso in cui il lavoratore è occupato e profonde un *effort* positivo, cioè si impegna; il dividendo è rappresentato da $W - e$, mentre b indica la probabilità, ipotizzata esogena, di diventare disoccupato, moltiplicata per la perdita in conto capitale.

La (3.16) riguarda il caso in cui il lavoratore è occupato ma fa *shirking*, cioè non si impegna: in questo caso, essendo $e=0$, il dividendo sarà W e la perdita attesa sarà la stessa del caso di *no-shirking*, ma tenendo conto anche del fatto che il lavoratore rischia di essere sorpreso a non impegnarsi.

Oltre alla probabilità esogena b occorre considerare, infatti, anche la probabilità (endogena) q di “essere colto in fallo” che dipende dal sistema di monitoraggio dell'impresa.

Risulta quindi evidente che il lavoratore si impegnerà se $V_e^{ns} \geq V_e^s$ (*non-shirking condition*); l'obiettivo è trovare il valore *minimo* di W che soddisfa la *non-*

shirking condition; tanto maggiore è W , infatti, tanto maggiore è la probabilità che la condizione sarà verificata come maggiore.

Dalla massimizzazione dei profitti si ottiene che la *non-shirking condition* verrà verificata come uguaglianza, quindi, in equilibrio avremo $Ve^{ns} = Ve^s = Ve$, di conseguenza varrà anche l'uguaglianza tra (3.15) e la (3.16), da cui risulta

$$W - e - b (Ve^{ns} - Vu) = W - (b+q) (Ve^s - Vu)$$

$$Ve - Vu = \frac{e}{q} > 0 \quad (3.17)$$

dove $(Ve - Vu)$ indica la differenza tra l'*asset value* del lavoro nel caso di occupazione (con *no shirking*) e di disoccupazione; il fatto che la (3.17) sia positiva, indica che i disoccupati preferiscono strettamente la situazione di occupati (cioè che, in equilibrio, vi è disoccupazione involontaria). Sottraendo rVe da rVu , avremo

$$rVe - rVu = W - e - b (Ve - Vu) - \underline{u} - a(Ve - Vu)$$

$$W^* = e + u + (a + b + r) \frac{e}{q} \quad (3.18)$$

Un volta trovata la (3.18) è possibile fare alcune osservazioni:

- se il costo dell'impegno è alto, le imprese devono pagare un salario ottimo maggiore, così come se l'utilità di riserva è alta, altrimenti per i lavoratori il licenziamento risulterebbe meno costoso;
- se la probabilità a di trovare un lavoro è alta, allora la remunerazione pagata dalle imprese sarà maggiore;
- se la probabilità b (di diventare disoccupato) è alta, i lavoratori sono scoraggiati a approfondire un livello elevato di impegno, pertanto avremo un W^* più alto per scoraggiare lo *shirking*²⁵;

²⁵ Questo avviene in quanto b è esogena, non rientra, cioè, nella sfera d'azione dell'impresa.

- se non c'è *turnover*, e quindi a, b sono nulli, i disoccupati non verranno mai assunti²⁶; di conseguenza la *non-shirking condition* sarà più piatta, in quanto indipendente dal livello di occupazione.

La (3.18) diviene $W^* = e + \underline{u} + r(e/q)$, ne consegue che W^* si abbassa;

- se r è alto, avremo un minore fattore di sconto, quindi verrà dato un minor “peso” ai *payoff* futuri e quindi i salari devono essere più alti;
- se la probabilità q , di essere scoperti durante un comportamento sleale, è alta, non c'è motivo di incentivare i lavoratori con elevati salari: nel caso limite (teorico) in cui q è infinito, cioè vi un perfetto sistema di monitoraggio, avremo che $W^* = e + u$, pertanto non sarebbe più necessario incentivare e non avremo più disoccupazione involontaria.

A questo punto, passiamo ad “endogeneizzare” la probabilità a di trovare lavoro²⁷. Indichiamo con L_f la forza lavoro, con N il numero delle imprese e con L il livello di occupazione scelta da ciascuna impresa, tale per cui $F'(eL) = W^*$.

Concentrandosi sull'equilibrio di *steady-state*, per cui il flusso istantaneo di lavoratori che passano dallo stato di disoccupato a quello di occupato è uguale al flusso istantaneo di lavoratori che passano da quello di occupato a quello di disoccupato, analiticamente avremo che

$$a(L_f - NL) = bNL$$

$$a = \frac{bNL}{L_f - NL} \quad (3.19)$$

sostituendo la (3.19) nella (3.18), avremo

$$W^* = e + \underline{u} + \left(\frac{bNL}{L_f - NL} + b + r \right) \frac{e}{q} \quad (3.20)$$

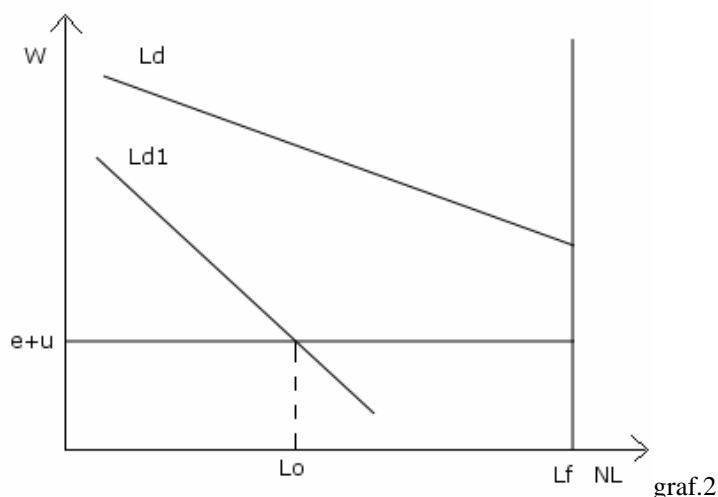
da cui

²⁶ *Ibid.*, p 455

²⁷ *Ibid.*, p 453

$$W^* = e + \underline{u} + \left(\frac{bL_f}{L_f - NL} + r \right) \frac{e}{q} \quad (3.21)$$

Passiamo ora ad analizzare l'offerta di lavoro: in un modello walrasiano senza problemi di incentivazione ed *effort* perfettamente contrattabile, al livello salariale $e + \underline{u}$, i lavoratori offrono in modo perfettamente elastico tutta la quantità di lavoro che le imprese domandano, mentre in L_f l'offerta è verticale; la domanda sarà negativamente inclinata in quanto espressione della produttività marginale del lavoro.



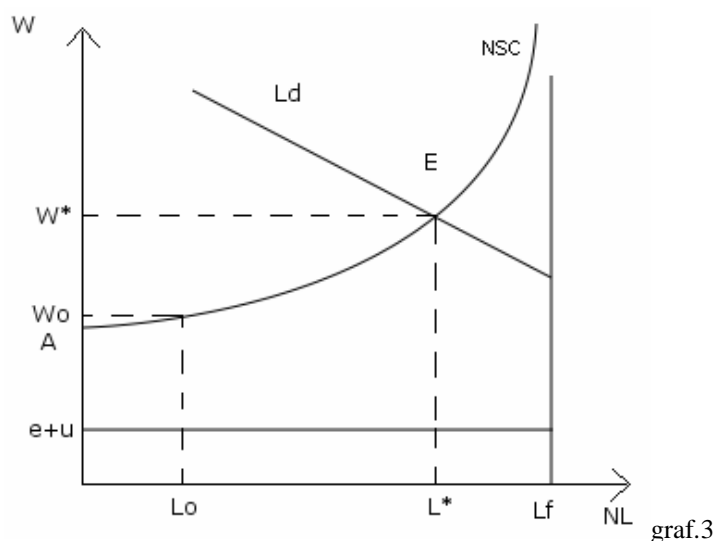
Inoltre, se per ipotesi, la produttività marginale del lavoro (nelle singole imprese ipotizzate identiche), in corrispondenza del livello d'occupazione L_f/N , moltiplicata per l'*effort* (cioè la produttività marginale in unità d'efficienza) è maggiore del livello di salario $e + \underline{u}$, cioè se vale

$$F'[e(L_f/N)] > e + \underline{u} \quad (3.22)$$

allora l'equilibrio (walrasiano) sarà al pieno impiego in L_f , poiché le imprese avrebbero incentivo ad assumere un numero di lavoratori maggiore della forza lavoro stessa (cosa che è tecnicamente impossibile).

Occorre inoltre notare che anche nel caso in cui la (3.22) non è verificata, tutta la disoccupazione che deriverebbe in caso di offerta orizzontale ($L_f - L_o$) è volontaria.

Abbandoniamo il modello walrasiano e rappresentiamo, all'interno di graf.2, la (3.20), cioè la NSC, che indica per ogni livello di occupazione il salario minimo per cui i lavoratori si impegnano



Il punto A ha valore $e + \underline{u} + (b + r)(e/q)$; se $NL = L_f$ allora avremo che il salario ottimale tende a infinito; questo, in altri termini, vuol dire che al livello di pieno impiego (L_f) è impossibile incentivare i lavoratori.

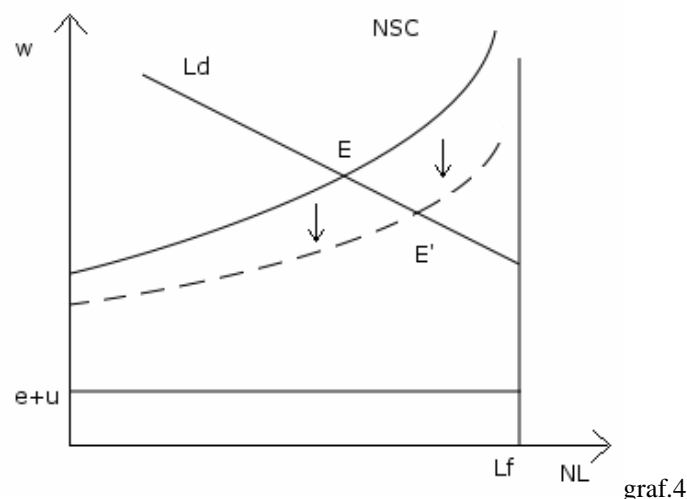
L'equilibrio si viene ad avere nel punto E , dato dall'intersezione tra la curva NSC e la curva di domanda: in E le imprese pagano W^* e assumono un livello di lavoratori pari a L^* . Il livello $(L_f - L^*)$ rappresenta disoccupazione involontaria e ovviamente alle imprese non conviene andare oltre L^* altrimenti i lavoratori non avrebbero incentivo ad impegnarsi.

Al di sotto della curva NSC i lavoratori non si impegnano, mentre al di sopra non avremmo la massimizzazione dei profitti dell'impresa.

La disoccupazione involontaria che deriva dal modello serve ad accrescere il costo del licenziamento: le imprese pagano un salario maggiore rispetto al costo opportunità del lavoratore, ma in caso di *shirking*, il lavoratore viene licenziato con probabilità q ; la disoccupazione involontaria viene quindi usata come strumento di disciplina.

Una variabile chiave nel modello è la probabilità q di monitoraggio, abbiamo già osservato che in presenza di un sistema di monitoraggio perfetto, non occorre incentivare i lavoratori.

Da un punto di vista grafico²⁸, all'aumentare di q avremo che la NSC si muove verso il basso



dove risulta evidente che la disoccupazione involontaria associata ad E è maggiore rispetto a quella associata a E', ottenuta a seguito dell'aumento di q .

Al limite, quando q tende a infinito, la NSC coincide con la funzione di offerta di lavoro del modello walrasiano.

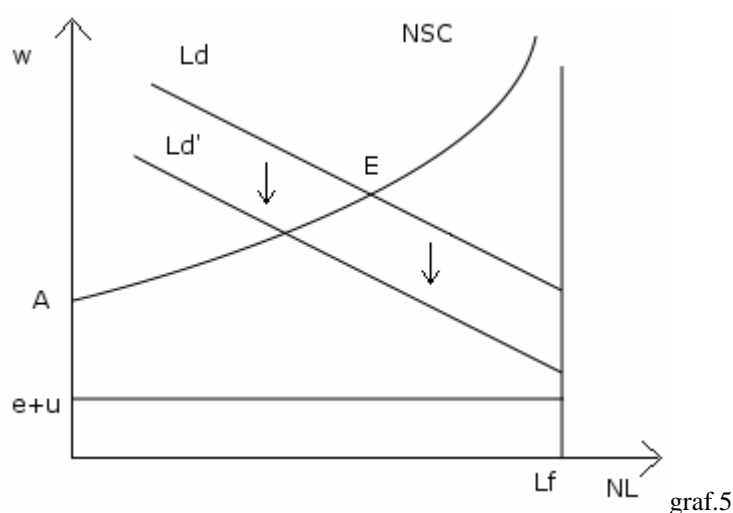
Il modello implica l'esistenza di un equilibrio caratterizzato da disoccupazione involontaria e suggerisce alcuni modi per poter influire su essa, tuttavia il modello così formulato è eccessivamente schematico, tale da non poter definire con precisione il livello di occupazione, nonostante possa definire a grandi linee le dinamiche e le relazioni tra le variabili che lo determinano.

²⁸ *Ibid.*, p 455

3.1.4 Considerazioni aggiuntive circa il modello di Shapiro
– Stiglitz: spinta della domanda, inclinazione della
NSC, efficienza e altre osservazioni

Ipotizziamo che a causa di fluttuazioni di breve periodo, si venga a determinare un calo della domanda di lavoro; a seguito di tale riduzione, l'equilibrio nel mercato del lavoro si sposterà verso il basso muovendosi sempre lungo la NSC.

Come risulta evidente da un punto di vista grafico (graf.5), il nuovo equilibrio è caratterizzato da un salario d'efficienza minore, al quale è associato un livello minore di occupazione.

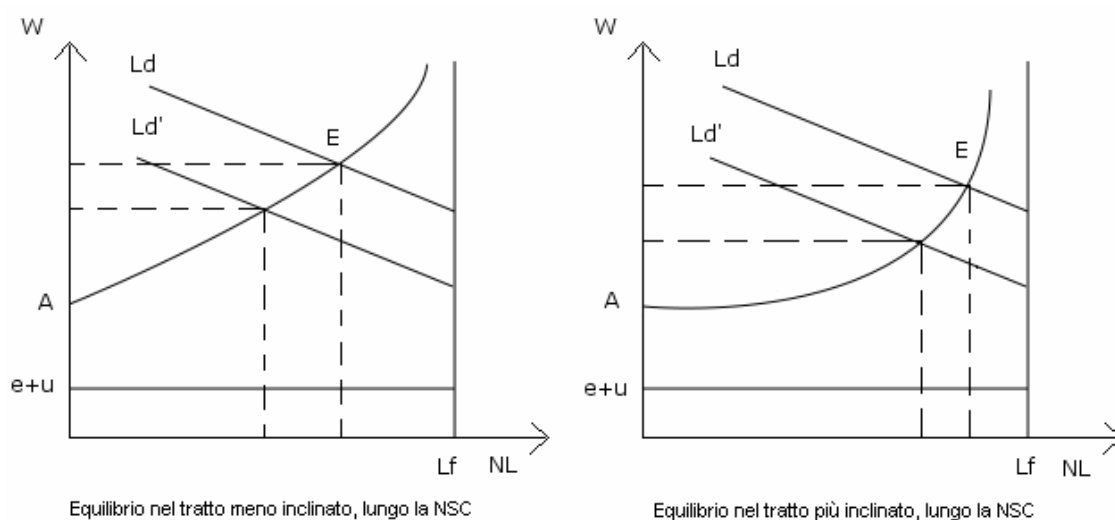


La disoccupazione aumenta e con essa anche il costo di licenziamento; si comprende, quindi, che per avere lo stesso costo di licenziamento della situazione precedente, è sufficiente un salario d'efficienza inferiore a prima.

Viceversa, con un aumento della domanda di lavoro, la riduzione della disoccupazione determina il calo del costo del licenziamento; quindi, nel caso di licenziamento, per un soggetto sarebbe più facile trovare un'occupazione alternativa. Le imprese "compenseranno" la riduzione del costo di licenziamento, causata dal calo della disoccupazione, con un aumento dei salari d'efficienza.

Al fine di valutare in modo completo, l'effetto di una variazione della domanda di lavoro, è opportuno sottolineare la rilevanza dell'inclinazione della NSC nel punto d'equilibrio.

Se il punto sulla NSC in cui si valuta la disoccupazione è troppo inclinato, il modello implica che una variazione della domanda avrà un impatto maggiore sui salari piuttosto che sull'occupazione (e viceversa nel caso in cui l'equilibrio sia su un tratto poco inclinato della NSC). Questa considerazione risulta evidente, in particolar modo, se analizzata dal punto di vista grafico



graf.6

da questa rappresentazione si vede chiaramente che tanto maggiore è l'inclinazione, nel punto d'equilibrio, della NSC tanto minore sarà l'effetto di una variazione di domanda sulla disoccupazione.

Più precisamente, se l'equilibrio si ottiene in un tratto molto inclinato della NSC, allora una variazione della domanda avrà effetti maggiori sul salario rispetto a quelli sull'occupazione; se, invece, l'equilibrio fosse su un tratto meno inclinato della NSC, allora gli effetti maggiori, a seguito di una variazione della domanda, si avrebbero sull'occupazione²⁹.

²⁹ *Ibid.*, p 457

Il modello, inoltre, implica l'inefficienza dell'equilibrio ottenuto: fino a che la produttività marginale del lavoro al pieno impiego è maggiore del costo dell'*effort* profuso dai lavoratori, non si ha una situazione di *first-best*.

Tale situazione si otterrebbe se tutti i lavoratori (L_f) fossero occupati e si impegnassero, ma d'altro canto l'autorità pubblica non può intervenire obbligando le imprese a giungere ad una situazione di piena occupazione; questo causerebbe l'incentivo a non impegnarsi e implicherebbe un livello di produzione nullo³⁰.

Tuttavia l'intervento statale non è sempre negativo: una politica di aumento della domanda di lavoro, infatti, è profittevole fino a quando il valore del prodotto in più, uguaglia il costo di produzione dell'unità aggiuntiva. In questo modo si ha un aumento del *welfare* complessivo. Il criterio con cui questo aumento di *welfare* viene diviso tra imprese e lavoratori dipende dal tipo di finanziamento dei sussidi.

Il modello di Shapiro e Stiglitz consente, inoltre, di fare altre importanti osservazioni³¹:

- l'evidenza empirica suggerisce che le imprese preferiscano licenziare piuttosto che segmentare il lavoro, nel momento in cui si renda necessaria una riduzione del personale. Con la riduzione delle ore lavoro, i dipendenti riuscirebbero a "catturare" una minor ammontare del *surplus* che essi stessi creano e quindi per prevenire comportamenti opportunistici, il salario ottimo diventa tanto maggiore quanto più elevata è la diversificazione del lavoro;
- ipotizziamo l'esistenza di un altro settore dove sono presenti procedure produttive che consentono un monitoraggio perfetto, dove quindi non

³⁰ *Ibid.*, p 457

³¹ *Ibid.*, p 458

viene generato *surplus* e questo si traduce in un maggior livello di giro d'affari. Nel settore con asimmetrie, invece, le imprese continuano a pagare più del salario d'equilibrio di *market clearing* e quindi spingono i lavoratori verso quelle procedure che sono monitorabili.

Introducendo lavoratori con diversi livelli di dedizione al lavoro, ovviamente sarà necessario un salario maggiore per i lavoratori meno "dediti"; le imprese dove le procedure necessitano di monitoraggio, non assumeranno lavoratori poco dediti che verranno quindi impiegato, in modo sproporzionato, con bassi salari in settori ad elevato *turnover*.

- se prima del licenziamento, la punizione per i comportamenti opportunistici consiste nell'abbassamento per un certo periodo, del salario, si viene ad avere un diverso impatto di breve periodo; la curva NSC sarà piatta e una variazione della domanda avrà effetti solo sull'occupazione;
- il modello di Shapiro e Stiglitz può essere ulteriormente esteso introducendo cauzioni sotto varie forme, tuttavia, un approccio di questo tipo, come vedremo in seguito, presenta sia lacune teoriche che la mancanza di evidenza empirica.

In conclusione è opportuno notare che il modello di Shapiro e Stiglitz non è generale, ma piuttosto "limitato" seppur estremamente esemplificativo: le variabili osservate non sono le sole che determinano il comportamento sia del lavoratore che dell'impresa. Pagare elevati salari può avere motivazioni diverse, così come la scelta di impegnarsi può derivare da processi decisionali che un approccio prettamente analitico come quello di Shapiro e Stiglitz non può considerare.

3.2 Osservazioni di sintesi sul modello di Shapiro e Stiglitz

La disoccupazione involontaria sembra essere una caratteristica persistente nei moderni sistemi economici; nell'ambito del modello neoclassico, dove si raggiunge sempre l'equilibrio in un contesto di perfetta concorrenzialità, i lavoratori non avranno incentivo a impegnarsi in quanto lo *shirking* è l'unica azione economicamente conveniente: anche se questi lavoratori venissero licenziati, troverebbero subito un'altra occupazione.

L'idea chiave della teoria dei salari d'efficienza, consiste nel rendere costoso il licenziamento; se un'impresa paga un salario maggiore rispetto a quello d'equilibrio, il lavoratore non ha incentivo a oziare; però, se ogni impresa adotta questa tattica pagando un salario maggiore di quello d'equilibrio, l'incentivo a non oziare sembrerebbe scomparire. In questo caso, però, il salario pagato sarebbe maggiore di quello d'equilibrio generalizzato, e quindi avremmo una diminuzione della domanda di lavoro e pertanto disoccupazione: il licenziamento torna ad essere costoso poiché i lavoratori licenziati avrebbero problemi a trovare una nuova occupazione. Ad ogni modo, la disoccupazione rappresenta il “meccanismo di disciplina dei lavoratori” e permette l'*enforcement* contrattuale³² mediante il rischio di perdite del flusso di quasi rendite positive associate alla condizione di occupato³³.

La teoria neoclassica stabilisce che una politica di alti salari determina soltanto un aumento dei costi e quindi un calo nei profitti; questa idea è stata abbandonata all'interno della teoria dei salari d'efficienza dove un aumento del

³² Nonostante l'assenza di misure oggettive delle performance e, quindi, nonostante l'impossibilità di avere l'*enforcement* contrattuale attraverso una corte.

³³ Scoppa V., *Contratti incompleti ed enforcement endogeno, Una rassegna della letteratura*, Economia Politica, p 404. 2003

salario, pur determinando un aumento dei costi, può generare, anche, un incremento dell'efficienza produttiva³⁴ e quindi dei ricavi.

Oltre alla funzione incentivante all'interno dell'impresa, quindi, i salari d'efficienza hanno anche un'importante funzione allocativa nel mercato del lavoro: uno dei messaggi chiave di questa teoria è appunto che il salario, svolgendo un ruolo incentivante, non può allo stesso tempo essere una variabile equilibratrice del mercato del lavoro.³⁵

Per ottenere i risultati precedentemente esposti occorre che i lavoratori abbiano una conoscenza perfetta³⁶ delle alternative di lavoro e che non ci siano posti vacanti. Inoltre l'ipotesi secondo la quale tutti i lavoratori sono identici, trova la sua giustificazione nell'obiettivo stesso dell'analisi: ipotizzando lavoratori diversi si introduce l'ulteriore problema della selezione avversa, problema che ai fini dell'analisi (originaria) di Shapiro e Stiglitz non viene considerato.

Il modello ha anche altre importanti implicazioni, che è opportuno segnalare per motivi di completezza:

- la presenza di sussidi determina un aumento del tasso di disoccupazione d'equilibrio in quanto disincentiva i disoccupati a ricercare un lavoro; il costo del licenziamento diminuisce e quindi per aver garantito l'impegno dei lavoratori è necessario un salario d'efficienza maggiore;
- i salari non rispondono in modo immediato a seguito di shock;
- l'equilibrio di mercato non è ottimo-paretiano; esistono casi in cui i sussidi sono utili così come l'intervento statale;
- il modello è coerente con la presenza di un tasso di disoccupazione naturale relativamente alto;
- i salari sono usati sia per allocare lavoro che per incentivare i lavoratori, quindi se le imprese non sono identiche e se per qualche impresa il costo

³⁴ Brucchi L., *op. cit.*, p 330

³⁵ *Ibid.*, p 316

³⁶ Pur essendo in un contesto di asimmetria informativa riguardante la relazione lavorativa.

dello *shirking* è maggiore rispetto alle altre, allora si spiega la persistenza nella dispersione del salario.

Occorre, inoltre, ribadire nuovamente che il modello di Shapiro e Stiglitz non è nato allo scopo di risolvere il problema del *moral hazard* in se, ma è sorto come teoria di microfondazione della disoccupazione involontaria persistente: questa considerazione è ovvia alla luce del contesto teorico in cui questo modello è stato formulato, tuttavia è anche necessaria per capirne i limiti.

In questa ottica, l'approccio dei due economisti potrebbe sembrare sviluppato ad hoc (e in realtà, di fatto, lo è) per spiegare il fenomeno e quindi, non necessariamente per rappresentare in modo appropriato gli aspetti microeconomici fine a se stessi.

Con quanto appena detto, però, non si vuole in alcun modo sminuire il contributo di Shapiro e Stiglitz, ma si vuole solo sottolineare che in realtà la disoccupazione involontaria potrebbe benissimo avere altre spiegazioni ed origini. E' possibile, infatti, sviluppare approcci microeconomici incentivanti che non implicano disoccupazione involontaria.

E' in questa ottica che ha preso vita il sistema delle cauzioni, sviluppato poi in quello dei profili salariali crescenti con l'anzianità, che ha dato impulso alla ricerca nel tentativo di ottenere schemi incentivanti che non implicassero disoccupazione involontaria. E' in questo ambito che si inserisce il filone di ricerca dei contratti *self-enforcing* e, in parte, quello dei tornei.

E' opportuno spendere qualche parola anche sulla dimensione sociale del lavoro e sul rapporto che questa potrebbe avere con la teoria dei salari d'efficienza.

Da un punto di vista sociale, infatti, la minaccia di licenziamento (così come qualsiasi altra minaccia) in caso di comportamenti opportunistici può avere effetti contrari a quelli desiderati.

Il modello dei salari d'efficienza, di fatto, si basa, come del resto quasi tutta la letteratura a riguardo, sulla contrapposizione (forse troppo netta) tra l'impresa e il lavoro; una politica salariale come quella descritta da Shapiro e Stiglitz, rischia di incrinare il rapporto tra le parti e impedisce l'identificazione del lavoratore nell'impresa stessa³⁷.

Inoltre, empiricamente si è visto che la punizione tramite il licenziamento viene, per lo più, usata nel caso di lavoratori temporanei e scarsamente produttivi; negli altri casi, infatti, sembrerebbe (da un punto di vista sia sociale che economico³⁸) preferibile intervenire in altri modi e ricorrere al licenziamento solo in casi estremi.

Questa problematica, di fatto, verrà meno nel filone dei contratti impliciti, dove il rapporto economico trae la propria forza dalla reputazione che le parti si sono costruite e dalla fiducia (intesa in termini economici) tra impresa e lavoratore.

Va infine sottolineato che il modello di Shapiro e Stiglitz risulta difficilmente applicabile in un contesto dove vi è una forte presenza sindacale, dove, cioè, la minaccia di licenziamento è meno credibile rispetto ad un contesto dove non vi è potere di mercato dal lato dell'offerta.

³⁷ Brucchi L., *op. cit.*, p 327

³⁸ Ad esempio, licenziare un lavoratore altamente produttivo, per una "mancanza" (rimediabile) commessa una sola volta, può essere un grosso errore dal punto di vista economico.

3.3 Il sistema delle cauzioni (*performance bonding*) e i profili salariali legati all'anzianità

Il sistema del *performance bonding*, pur avendo avuto poco seguito sia a livello teorico che empirico, ha rappresentato per la letteratura scientifica successiva ai salari d'efficienza, un importante punto di riferimento per quanto riguarda, in particolare, i risultati teorici ottenuti. Il sistema delle cauzioni nasce a partire dai salari d'efficienza, uno dei primi articoli in cui compaiono è, infatti, il commento di Carmichael³⁹ al modello di Shapiro – Stiglitz.

Come vedremo nei capitoli successivi, il sistema delle cauzioni è stato usato come punto di riferimento per la costruzione di altri sistemi di incentivazione, in particolare il lavoro di Lin e Yang⁴⁰ (riguardante il sistema dei tornei e i loro effetti macroeconomici) si pone come obiettivo la costruzione di un sistema di incentivazione avente gli stessi risultati teorici del modello delle cauzioni senza però essere affetto dai problemi intrinseci del modello stesso.

Risulta pertanto evidente la necessità di chiarire, seppur sommariamente, i punti salienti di questa impostazione ed in particolare le cause del suo “fallimento”; passiamo, quindi, a introdurre il problema delle cauzioni.

La cauzione, o caparra, è un pegno che il lavoratore paga all'impresa nel momento in cui viene assunto; la cauzione, che inizialmente possiamo vedere come un certo ammontare monetario, è usata a scopi deterrenti per evitare che i lavoratori si comportino in modo opportunistico. L'impresa oltre al licenziamento, minaccia di non restituire la caparra: in una situazione di equilibrio nel mercato del lavoro, dove cioè non vi è disoccupazione involontaria, il salario che i lavoratori percepiscono dall'impresa è pari (in assenza di sistemi

³⁹ Carmichael H.L., *Can Unemployment Be Involuntary? Comment*, American Economic Association, vol. 75(5), pp 1213-1214. 1985

⁴⁰ Lin C., Yang C., *Tournaments as a worker discipline device*, Economic Inquiry 39, pp 644-650. 2001

di incentivazione) al salario alternativo e, quindi, per i lavoratori è sempre conveniente non impegnarsi in quanto il costo del licenziamento è praticamente nullo.

Come abbiamo parzialmente già visto, l'incentivazione può derivare dal rendere costoso il licenziamento: l'aumento del costo di licenziamento può esser fatto in vari modi, molti dei quali prevedono il pagamento di un salario maggiore rispetto a quello d'equilibrio, generando, in questo modo, disoccupazione involontaria; di fatto questa è l'idea principale di fondo del modello di Shapiro–Stiglitz.

Con il pagamento di una cauzione, le imprese possono ottenere l'impegno dei lavoratori senza per questo dover pagare un salario maggiore rispetto a quello fissato dall'equilibrio di mercato. Con il sistema delle cauzioni, si elimina, almeno al livello teorico, il problema della disoccupazione involontaria.

Inoltre, facendo pagare ai lavoratori una cauzione (pari agli extra salari che si attendono di ottenere in futuro), l'impresa riduce i costi legati al lavoro: in equilibrio, le imprese impiegheranno i lavoratori fino al punto in cui l'ultimo lavoratore assunto (lavoratore marginale) percepisce un salario pari alla propria produttività; tutti i lavoratori, quindi, risulteranno avere una produttività pari a quella che avrebbero se fossero impiegati da qualsiasi altra parte.

Il sistema delle cauzioni, quindi, presenta notevoli pregi teorici, in termini di effetti macroeconomici, tuttavia l'evidenza empirica ha mostrato come questo tipo di incentivi sia quasi del tutto assente, quantomeno nella formulazione originaria.

Anche a livello teorico, ci devono essere dei presupposti che non necessariamente valgono nella realtà; in particolare il pagamento di un certo ammontare all'inizio della vita lavorativa, presuppone la possibilità di accesso ad un sistema di credito: è, tuttavia, impossibile che un lavoratore che si appresta

ad entrare nel mondo del lavoro, presumibilmente senza risparmi né capitali, possa pagare una cauzione senza accedere al sistema di credito; tale accesso, però, non è facile, in particolare per soggetti senza garanzie.

3.3.1 Profili salariali legati all'anzianità: i lavoratori come "creditori" dell'impresa

La teoria economica ha cercato di superare questo problema, introducendo il sistema delle cauzioni sotto la forma di profili salariali crescenti nel tempo.

Uno dei primi contributi teorici relativi ai profili salariali crescenti con l'anzianità è di Lazear⁴¹; nel modello viene ipotizzato che l'impresa operi in un contesto di perfetta concorrenza ed utilizzi, nel processo produttivo, solo il fattore lavoro; viene inoltre ipotizzato che ogni lavoratore abbia solo una mansione (viene escluso il problema del *multitasking*). Ciascun lavoratore può approfondire un livello di *effort* alto o basso che genera un livello di prodotto individuale (q) rispettivamente alto o basso.

La permanenza all'interno della realtà industriale determina un aumento delle competenze e abilità del lavoratore e pertanto il valore del prodotto individuale risulta essere funzione positiva, con rendimenti decrescenti, del tempo, ossia $q(t)$ con $q' > 0$ e $q'' < 0$.

$A(t)$ indica l'utilità di riserva anch'essa crescente nel tempo, che possiamo vedere come la funzione che descrive il valore del tempo libero; l'efficienza interna all'impresa richiede, inoltre, che il livello di *effort* (non verificabile ma monitorabile solo occasionalmente) da parte del lavoratore sia alto.

⁴¹ Lazear, E. P. (1979), Why is there mandatory retirement, *Journal of Political Economy*, 87, 1261-64.

Se il profilo salariale è crescente rispetto al tempo, allora l'incentivo induce a impegnarsi: supponiamo che lo schema retributivo preveda il medesimo valore attuale di $q(t)$ ma che la retribuzione sia al di sopra di $q(t)$ per valori molto alti di t e al di sotto di $q(t)$ per valori bassi di t . In questo modo, gli incrementi salariali sono condizionati soltanto all'anzianità e il lavoratore ha incentivo a approfondire il livello efficiente di *effort*; uno schema di questo tipo, tuttavia, per mantenere l'efficienza, ha bisogno di un sistema di pensionamento obbligatorio in quanto non è nell'interesse del lavoratore lasciare l'impresa al tempo socialmente efficiente.

Ripercorriamo concettualmente l'analisi svolta da Lazear, semplificandone alcuni punti e mostrandone l'idea di fondo: l'evidenza empirica è consistente con questa teoria e suggerisce che i salari crescono nel corso della carriera.

Ipotizziamo che le *performance* non siano osservabili e che ci sia un contratto implicito⁴², cioè un accordo non legalmente vincolante tra impresa e lavoratore, che stabilisce che il lavoratore fino ad un certo tempo t riceverà un salario minore della produttività, superato t riceverà un salario superiore alla produttività; lo schema è accompagnato dalla minaccia di licenziamento nel caso il lavoratore sia sorpreso ad non impegnarsi.

L'aumento del salario avviene in modo automatico in base all'anzianità di servizio. E' come se il lavoratore concedesse un prestito all'impresa, impresa che farà in modo tale che i salari percepiti complessivamente nell'arco della vita lavorativa (in termini attesi) siano uguali alla produttività.

In questo modo, in una prospettiva pluriperiodale (l'intera vita lavorativa del dipendente), l'impresa non ha bisogno di pagare salari maggiori della produttività: i lavoratori non avranno convenienza a fare *shirking* semplicemente perché interessati a "riavere ciò che hanno prestato all'impresa".

Questo schema contrattuale è del tutto equivalente al sistema delle cauzioni precedentemente descritto, la differenza sta nel fatto che la cauzione è

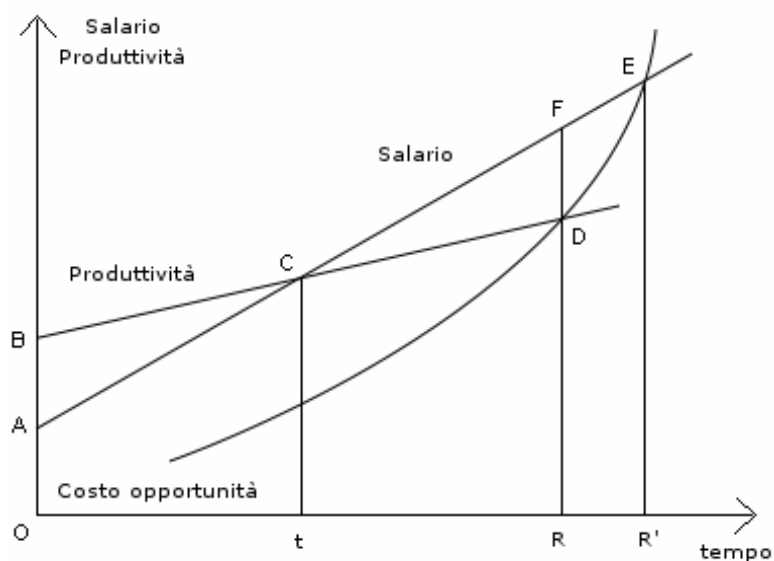
⁴² Il concetto di contratto implicito verrà approfondito nel capitolo seguente.

“distribuita” sull’intera vita lavorativa e non è richiesto un esborso monetario da parte del lavoratore all’inizio del rapporto di lavoro.

Questo approccio, quindi, funziona se i lavoratori si aspettano di restare presso l’impresa per lungo tempo e nonostante la presenza di vincoli di liquidità.

Le imprese sarebbero incentivate a impiegare i lavoratori fino al tempo t (graf.7) per poi licenziarli; per le imprese sarebbe, cioè, conveniente licenziare i dipendenti non appena la remunerazione raggiunge la produttività. Si rende, quindi, necessario un meccanismo di reputazione dell’impresa affinché i lavoratori siano disposti ad aderire al contratto implicito.

I profili salariali crescenti con l’anzianità, come già osservato, richiedono un sistema di pensionamento obbligatorio. Il tempo socialmente efficiente di pensionamento si viene ad avere nel momento in cui la produttività marginale del lavoro è uguale al suo costo opportunità (visto come disutilità marginale dell’impegno). Definita la situazione efficiente, il sistema retributivo viene strutturato in modo che il valore atteso dei salari, ricevuti nel corso della vita lavorativa, sia pari al monte salario totale che un lavoratore avrebbe ricevuto se fosse stato retribuito in base alla produttività. Da un punto di vista grafico, una volta trovato il punto D (graf.7), la retribuzione viene strutturata in modo che le aree $OAFR$ e $OBDR$ siano uguali



graf.7

Nel grafico⁴³ sono rappresentate la produttività del lavoratore, il salario e il costo opportunità; il periodo socialmente efficiente di pensionamento è R , ma poiché i lavoratori a partire da t ottengono un salario maggiore della produttività, potrebbero decidere di andare in pensione successivamente, ad esempio in R' ; occorre quindi il pensionamento obbligatorio o un sistema di incentivi⁴⁴.

L'evidenza empirica suggerisce che gli incrementi salariali dovuti all'istruzione e agli anni di esperienza sono dovuti all'aumento del capitale umano generico, mentre la correlazione positiva tra salari e anzianità è dovuta all'aumento del capitale umano specifico; secondo altri economisti, invece, tale correlazione positiva può dipendere anche dalla natura incentivante dei salari progressivi.

L'impostazione basata sul capitale umano prevede che il lavoratore diventi più produttivo con l'anzianità, mentre l'impostazione basata su asimmetrie informative non si esprime in merito.

L'evidenza empirica suggerisce che la produttività rimanga la stessa, se non addirittura cali, con il passare del tempo.

I lavoratori anziani, quindi, non guadagnano più dei giovani poiché maggiormente produttivi, ma semplicemente perché sono "creditori" dell'impresa.

L'evidenza empirica dice, inoltre, che nei settori dove è difficile monitorare le *performance*, i salari sono crescenti; là dove esiste un efficace sistema di monitoraggio, i salari sono, tendenzialmente, legati alle *performance*.

E' stato inoltre riscontrato che i salari dei lavoratori autonomi sono "piatti" (rispecchiamo quindi la produttività) mentre quelli dei subordinati sono crescenti. Inoltre a parità di età, lavoratori assunti di recente percepiscono un salario più basso.

⁴³ Nicita A., Scoppa V., *op. cit.*, p 136

⁴⁴ Congiuntamente ad un sistema reputazionale che impedisca all'impresa di licenziare in t .

Facciamo un passo indietro: entrambe le parti, nel decidere, considerano il futuro ponendosi mentalmente in quella situazione, occorre pertanto utilizzare la *backwards induction*.

Un salario fisso non è un deterrente, l'unico modo per l'impresa è porre il lavoratore in condizione tale che fino all'ultimo periodo ciò che riceve come remunerazione è maggiore del beneficio derivante dal non impegnarsi; questo si può ottenere, come già visto, con un profilo salariale crescente.

Da dove deriva il *surplus* necessario per garantire l'*enforcement*? In questo caso la fonte di *surplus*, necessario per l'*enforcement* è la reputazione dell'impresa: i lavoratori si basano sul presente e sul passato per valutare la veridicità delle promesse dell'impresa; nel momento in cui le aspettative di onestà sono disattese, sarà difficile per l'impresa trovare nuovi lavoratori.

Pertanto, se l'impresa tiene alla propria reputazione saranno necessarie delle quasi rendite.

Se l'impresa, infatti, disattendesse le promesse, questo si tradurrebbe in richieste salariali maggiori da parte dei potenziali partner, con conseguente riduzione dei profitti.

Nonostante le imprese non siano incentivate a comportarsi in modo opportunistico, tuttavia, possono realizzarsi casi in cui sia la realtà economica (come ad esempio fasi di crisi) a determinare l'impossibilità di mantenere gli impegni presi; si comprende, quindi, il motivo per cui questo schema di retribuzione sia più frequente nelle imprese solide e a bassa rischiosità.

4. Incompletezza contrattuale, *enforcement* endogeno e i contratti *self – enforcing*: teoria dei contratti impliciti

Ogni relazione economica, anche la più semplice e basilare è regolata da un contratto, un accordo definito dalle parti coinvolte le quali, in un contesto di asimmetria informativa, sono “soggette” a tentazioni opportunistiche, stemperate da un adeguato sistema di *enforcing*, che spesso, ma non sempre, è garantito da un'autorità esterna.

La costruzione e la formalizzazione dei contratti impliciti, cioè di accordi informali, si è resa necessaria a partire dall'impossibilità (e/o dalla “non volontà”) di definire contrattualmente tutti i possibili aspetti che rilevano nel rapporto lavorativo (o in generale in qualsiasi rapporto economico); questo ha avuto come conseguenza l'inefficacia dell'*enforcement* basato su una corte.

La teoria dei contratti impliciti (quantomeno in una prima versione) rappresenta un'altra importante impostazione teorica (nata anch'essa nell'ambito della NMK) volta a microfondare le rigidità salariali, a partire da un contesto microeconomico con decisioni ottimizzanti da parte dei soggetti coinvolti (lavoratori e imprese).

I contratti impliciti intesi in questo senso, però, appartengono ad una “prima generazione” teorica e non sono quelli a cui si fa generalmente riferimento nella letteratura più recente.

Al fine di evitare fraintendimenti, infatti, occorre osservare preliminarmente che esistono “due” approcci ai contratti impliciti: nei primi modelli il problema affrontato è nuovamente quello dell'allocazione efficiente del rischio, risolto

tramite un salario fisso, attraverso il quale si ottiene una giustificazione della vischiosità salariale.

In questa prima impostazione, i lavoratori sono “tutelati” dalle oscillazioni del prodotto: vi è, infatti, una componente “assicurativa” che impedisce alle imprese di variare il salario o licenziare in base alle fluttuazioni nel mercato del lavoro¹.

La remunerazione del lavoratore, infatti, può essere pensata² pari alla produttività marginale meno il premio assicurativo pagato all’impresa in caso di fluttuazioni positive oppure, più una indennità di rischio nei momenti in cui la domanda di lavoro è bassa³. Questa impostazione, però, pur prendendo lo stesso nome, non è “vicina” a quella dei contratti *self-enforcing*, dove, invece, il problema centrale è l’opportunismo post contrattuale (*moral hazard*).

I contratti impliciti della “seconda generazione” prendono anch’essi vita dalle difficoltà pratiche nella costruzione di un contratto esplicito, tuttavia sono volti alla risoluzione di problematiche profondamente diverse e, quindi, partono da presupposti differenti.

Nei contratti impliciti, intesi nel senso di contratti *self-enforcing*, non vi è, infatti, nessuna problematica di allocazione del rischio (i soggetti sono ipotizzati neutrali) ma nasce il problema della credibilità delle parti e, in particolare nel caso della promessa di pagamento di un bonus, nasce il problema del doppio *moral hazard*.

L’obiettivo finale di questo lavoro, consiste nell’analizzare gli effetti occupazionali degli schemi di incentivazione; si comprende, quindi, il motivo per cui la “prima generazione” di contratti impliciti (in modo del tutto simile al modello d’agenzia) non rappresenti un interessante spunto d’analisi in tal senso.

¹ Questo schema di remunerazione implica salari superiori alla produttività marginale in caso di fluttuazioni negative, e minori in caso di fluttuazioni positive.

² Salituro B., *La teoria dei contratti impliciti, la rigidità salariale e la disoccupazione involontaria: una sintesi dei risultati*, Economia Politica, p 41. 1985

³ Le imprese, quindi, non potranno, né vorranno, (se non in casi particolari e di profonda crisi) licenziare anche qualora vi fosse un calo della domanda di lavoro. Similmente i lavoratori tenderanno a restare all’interno delle realtà industriali con la quale hanno già instaurato un rapporto.

Gli aspetti che ho approfondito e portato avanti nei paragrafi seguenti si riferiscono, infatti, ai contratti impliciti nel senso *self – enforcing*.

Nell'analisi di questa nuova tipologia di contratti, alcuni elementi teorici, sul quale si concentravano le impostazioni precedenti, perdono consistenza (in particolare l'esogeneità dell'*enforcement*) e vengono sostituiti da nuovi concetti.

Occorre, quindi, (ri)partire quasi da zero, analizzando nuovamente gli aspetti microeconomici relativi alle scelte dei soggetti e ponendo una “rinnovata” attenzione al problema dell'*enforcing* contrattuale, che diventerà un importante (se non il principale) oggetto d'analisi.

Nei contratti d'agenzia, ad esempio, l'espedito di legare la remunerazione al prodotto, serve⁴ a definire un elemento oggettivo di riferimento osservabile (da tutti i soggetti, compresa una autorità esterna) e “rappresentativo”⁵ del lavoro individuale; in questo modo vi è qualcosa di oggettivo su cui basare le eventuali rimostranze.

Un discorso analogo vale, limitatamente, anche per i contratti che basano la remunerazione sui salari d'efficienza: una volta *determinato* il salario d'efficienza, infatti, è possibile definirlo contrattualmente in modo formale⁶ (consentendo in questo modo il ricorso ad una corte in caso di rimostranze). La definizione esplicita del salario d'efficienza, però, consente di eliminare le tentazioni opportunistiche del principale ma, di per sé, non garantisce a monte l'*enforcement* contrattuale. Nei salari d'efficienza, infatti, non esistono misure oggettive delle *performance* e, quindi, non è possibile assicurare l'*enforcement* attraverso un'autorità esterna. In quest'ottica i contratti *self-enforcing* e i salari d'efficienza si avvicinano in quanto la problematica da cui muovono è la

⁴ Nei casi di asimmetrie informativa con incertezza.

⁵ Nonostante i problemi rappresentati dalla componente stocastica.

⁶ Nicita A., Scoppa V., *op. cit.*, p 125

medesima: come ottenere l'impegno in assenza di una corte che renda vincolante il contratto?

I punti in comune, però, terminano immediatamente, dal momento che i salari d'efficienza, attraverso il pagamento (ad esempio, anticipato) di un "salario fisso", risolvono il *moral hazard* senza la "necessità", come vedremo nei contratti *self-enforcing*, di meccanismi reputazionali volti a eliminare l'opportunismo d'impresa.

Nelle prime impostazioni teoriche, l'esigenza fondamentale consisteva nel dare rigore formale ad un problema fino ad allora non analizzato e, grazie alla natura esplicita e completa dei contratti, risultava del tutto "naturale" l'ipotesi di *enforcement* esogeno.

Tuttavia, il più delle volte è impossibile formulare *ex ante* un contratto esplicito e completo a causa dell'impossibilità di prevedere tutte le possibili situazioni future (nonché di attribuire probabilità valide a quelle prevedibili) ma anche a causa della frequente mancanza di misure oggettive delle *performance* che possono non essere verificabili da un'autorità esterna la quale può, in aggiunta, essere non competente in materia.

I contratti impliciti sono, pertanto, caratterizzati da un certo livello di *gap* contrattuale, colmato direttamente dalle azioni adottate dai soggetti coinvolti nel rapporto.

Questa nuova tipologia di contratti, oltre a rendersi necessaria a causa di problemi di natura oggettiva, potrebbe essere preferita direttamente dalle parti coinvolte nel momento in cui, ad esempio, gli indicatori oggettivi disponibili, a cui agganciare i termini contrattuali, risultino essere misure distorte delle *performance*.

Le imprese, in particolare, possono essere maggiormente propense a stipulare un contratto implicito nel caso di *multitasking* (il lavoratore ha diverse mansioni da svolgere): se per alcuni *task* è possibile definire incentivi espliciti, mentre per altri questo risulta difficile, il lavoratore tenderà a sacrificare le mansioni per le quali non vi è un incentivo esplicito a scapito, quindi, della redditività del datore di lavoro, che dipende anche da esse. Questo problema viene risolto prevedendo, per tutti i *task*, incentivi impliciti.

Ovviamente il problema dell'asimmetria informativa gioca un ruolo centrale in quanto rappresenta la stessa "ragion d'essere" di tutta la letteratura analizzata, tuttavia il punto di partenza dell'analisi è abbastanza diverso nei contratti impliciti rispetto agli altri modelli.

E' opportuno un chiarimento: nei contratti d'efficienza il nucleo dell'analisi consiste nell'individuare una remunerazione incentivante che "risolve il problema" del *moral hazard* tramite la creazione di disoccupazione⁷.

Nei contratti impliciti, invece, la remunerazione contingente, in sé, non è importante: in questa impostazione, infatti, non si determina una specificazione salariale, funzione dell'impianto parametrico su cui l'impresa ha (più o meno) potere.

La creazione di un contratto implicito è, di per sé, (apparentemente) facile; il problema, però, è lo spostamento dall'ottica di breve periodo a quella di lungo periodo; in sostanza, il problema riguarderà da ora in poi la capacità del contratto di essere vincolante e non tanto il contratto in sé.

⁷ Ovviamente, il discorso non è così semplice, ci sono moltissimi aspetti da considerare. I termini, a dir poco riduttivi, con il quale viene descritta la teoria dei salari d'efficienza servono solo a sottolineare l'idea che sta alla base dell'approccio.

I contratti impliciti, detto in altri termini, sono estremamente “flessibili” e “dinamici” (essendo per l'appunto accordi informali) e quindi, per loro stessa natura, sono relativamente facili da costruire.

Molto difficile, però, è fare in modo che le parti siano effettivamente incentivate a rispettarli, in quanto questi accordi, privi di valore legale, non sono altro che uno scambio di promesse e di aspettative condivise. Come rendere vincolante questa tipologia di contratti dal momento che una corte non può esprimersi su qualcosa non previsto “nero su bianco”?

Un elemento chiave del “nuovo” *enforcement* consiste nella possibilità di valutare *ex post* la correttezza dell'operato dell'altra parte⁸. In questo contesto, la teoria dei contratti impliciti assume un ruolo centrale e ha consentito di superare il problema della mancanza di *enforcement* esogeno.

Sono le parti che svolgono il ruolo di “autorità” in un modo “migliore” di quello che potrebbe fare una corte, in quanto dispongono di esperienza e informazioni tecniche che nessun giudice può avere.

I lavoratori, in particolare, cercano di mantenere una reputazione di affidabilità. La reputazione, come vedremo più in dettaglio, è un vero e proprio bene economico: con una buona reputazione è più facile mantenere ed ottenere nuove relazioni economiche; con una cattiva reputazione questo, invece, diventa difficile se non impossibile.

La reputazione, quindi, induce a rinunciare ai vantaggi di breve periodo, derivanti da azioni sleali, in favore di un più alto *pay-off* futuro; in altre parole, vi è la rottura del legame tra il salario corrente e le decisioni (microeconomiche) “contingenti” sull'offrire o meno lavoro (ma anche sull'impegnarsi o meno).

⁸ Scoppa V., *op. cit.*, p 397

La questione reputazionale, ovviamente, non si limita ai soli lavoratori ma si estende anche alle imprese; si comincia, quindi, ad intuire quali siano le basi del “nuovo” *enforcing* contrattuale.

Va infine sottolineato che il meccanismo reputazionale acquisisce forza e consistenza solo tramite la dimensione temporale: affinché si possa creare la reputazione è necessario (ma non sufficiente) che la relazione economica sia ripetuta nel tempo.

In questa prospettiva, lavoratori ed imprese sono spinti ad entrare in una relazione economica dove gli elementi chiave vengono definiti contrattualmente (anche se non in forma esplicita) per periodi di tempo prolungati e quindi indipendenti (in una certa misura) dalle fluttuazioni economiche⁹.

Si comincia a capire quali siano le principali linee guida del nuovo tipo *enforcing* contrattuale (che prende il nome di endogeno) basato sulle dinamiche interne al mercato in cui i soggetti, coinvolti nella relazione, operano.

L'*enforcement* endogeno si suddivide, a sua volta, in bilaterale (interno) e multilaterale (esterno)

Appartengono al primo tipo, quei meccanismi che assicurano l'applicazione del contratto tramite strategie di sanzionamento, come il licenziamento, adottate direttamente dalle due parti coinvolte nella relazione economica.

I datori di lavoro, ad esempio, licenzieranno i lavoratori che non si impegnano; le banche non concederanno credito alle imprese inaffidabili; i consumatori “puniranno” certi prodotti se non sono soddisfatti della qualità.

⁹ Salituro B., *op. cit.*, p 37

In particolare, nel caso del mercato del lavoro, l'*enforcement* interno è tanto più efficace quanto maggiore è il costo che l'agente deve sostenere in caso di licenziamento.

Una caratteristica fondamentale dell'*enforcement* endogeno interno è data dal fatto che non vi è la necessità che le informazioni riguardanti la relazione lavorativa, “escano” all'esterno della relazione stessa.

Il secondo tipo di *enforcement* endogeno è quello esterno; questo tipo di *enforcement* si rende necessario, in particolar modo, quando i costi di “migrazione” del soggetto licenziato sono bassi o addirittura nulli. L'obiettivo, anche in questo caso, è rendere difficoltoso, lo spostamento verso un altro partner in modo da rendere maggiormente costoso l'eventuale licenziamento.

La differenza fondamentale rispetto all'*enforcement* interno, sta nel fatto che in questo caso è necessario che le informazioni rilevanti circa la relazione economica vengano trasferite; il “maggior costo di licenziamento” si ha, infatti, grazie al coinvolgimento di soggetti esterni alla relazione.

In entrambi i caso, il problema dei contratti impliciti consiste nella credibilità delle “promesse” che le parti, pur di coinvolgersi a vicenda, fanno.

Supponiamo, ad esempio, che un'impresa, in caso di impegno da parte dei lavoratori, prometta, congiuntamente al rinnovo del contratto, il pagamento di un bonus.

In assenza di un meccanismo esogeno di *enforcement*, un contratto di questo tipo, ovviamente, è auto-vincolante se e solo se tutte le parti hanno incentivo a rispettarlo e questo può avvenire solo in un'ottica di lungo periodo e se vi è la creazione di *surplus*.

Per avere l'*enforcement* endogeno, quindi, può bastare un certo livello di *surplus*, che permette alle parti di ottenere quasi rendite positive.

Se non ci fosse la creazione di *surplus* all'interno della relazione, infatti, entrambe le parti potrebbero comportarsi in modo opportunistico, anche se questo non è sempre vero.

Ovviamente l'esistenza di quasi rendite è in contrasto con l'idea di mercati perfettamente concorrenziali in quanto i soggetti possono disporre, a seguito dell'instaurazione della relazione economica, di un valore maggiore rispetto alle alternative di mercato.

Perché l'esistenza di *surplus* interno alla relazione non è condizione necessaria per avere *enforcement* endogeno? Per rispondere a questa domanda è necessario tornare alla distinzione tra *enforcement* endogeno interno ed esterno: le considerazioni fino ad ora fatte circa il *surplus* si rivolgono, essenzialmente, al primo tipo di *enforcing*.

Come vedremo, infatti, l'*enforcement* esterno si può avere (ed è questo uno¹⁰ dei motivi per cui è stato introdotto) anche in assenza di *surplus* generato all'interno della relazione; in questo caso il costo del licenziamento sembrerebbe, però, nullo.

Quanto appena detto, però, non implica che non vi possa essere *surplus* anche con *enforcement* multilaterale: quasi rendite positive, infatti, possono essere presenti, ma non nascono all'interno del rapporto bensì dalla permanenza prolungata nel mercato o, ad esempio, dall'adesione a istituzioni che operano sul mercato.

In ogni caso, risulta evidente che il *surplus* generato all'interno della relazione economica in realtà, altro non è che il "valore di mercato" della relazione economica in se. Le parti, quindi, non hanno come scopo la semplice reiterazione

¹⁰ Un altro motivo all'origine dell'*enforcement* endogeno esterno, riguarda la durata, tendenzialmente limitata e nota, della relazione. Come vedremo più in dettaglio nel paragrafo successivo, una condizione necessaria per avere *enforcement* interno è che la relazione abbia durata incerta.

del rapporto lavorativo, bensì il mantenimento del valore che essi stessi contribuiscono a generare.

Risulta chiaro, quindi, che un altro problema centrale nell'analisi dei contratti impliciti consiste nello studio delle fonti del *surplus*. In estrema sintesi, possiamo definire (in ogni caso) la minaccia di sanzionamento come la minaccia di far cessare i flussi delle quasi rendite dovute alla creazione di *surplus* (sia nel caso che questo si determini all'interno della relazione lavorativa, sia nel caso in cui il *surplus* derivi dalla permanenza dei soggetti sui mercati).

Ovviamente, anche la situazione del mercato del lavoro gioca un ruolo fondamentale: se vi è un eccesso di domanda di lavoro, infatti, qualora le imprese promettessero dei bonus, i lavoratori sarebbero giustificati a credere a tale promessa, in quanto il costo della loro "uscita", sarebbe molto elevato per l'impresa. Con alta disoccupazione, le imprese terranno fermi i salari e i lavoratori saranno motivati dalla minaccia del licenziamento accresciuta dalla difficoltà di trovare un nuovo posto data la situazione economica.

Occorre inoltre sottolineare che non è il salario che gioca un ruolo nel *surplus*, ma è la differenza tra il salario percepito e quello alternativo.

Passiamo ad analizzare in modo più accurato il problema dell'*enforcement* contrattuale, definendo in maggior dettaglio gli aspetti fondamentali di quello endogeno e introducendo le sue due principali categorie.

4.1 Contratti *self – enforcing*: dall'esogeneità all'endogeneità dell'*enforcement* contrattuale

Ciascun individuo, coinvolto in un contratto, tende per propria natura ad essere non solo egoista ma anche opportunist (a scapito della controparte). Se una corte non è in grado di far rispettare un contratto, allora tale compito “spetta” al mercato.

Al suo interno, infatti, il mercato è in grado di creare sistemi di *enforcing* endogeni (che si viene ad identificare con molti¹¹ concetti, in particolare con quello di reputazione).

I contratti stabiliti in questa situazione si dicono *self-enforcing* in quanto il loro rispetto è garantito dall'interesse e dall'autonomo comportamento dei soggetti stessi¹².

L'esistenza di *enforcement* endogeno, però, è subordinata ad alcuni requisiti fondamentali che ne rappresentano le condizioni necessarie¹³; in altri termini, per avere *enforcement* endogeno occorre che

- la relazione economica sia ripetuta nel tempo e, qualora il contesto sia bilaterale, produca *surplus* consentendo, quindi, quasi rendite positive (che rappresentano il maggior valore che la relazione economica possiede, rispetto alle alternative a cui le parti possono rivolgersi);
- le parti incaricate del sanzionamento ricevano le informazioni rilevanti circa la relazione economico e abbiano strumenti idonei a sanzionare eventuali comportamenti inidonei; occorre, quindi, che i comportamenti delle parti siano osservabili. Ovviamente, l'identificazione dei soggetti che

¹¹ All'interno dell'*enforcement* endogeno, non vi è solo la reputazione, ma sicuramente ne rappresenta uno degli aspetti preponderanti.

¹² Scoppa V., *op. cit.*, p 392

¹³ *Ibid.*, p 392

hanno il compito di sanzionare, è diversa, come vedremo immediatamente, sulla base della tipologia di *enforcing* endogeno;

- la correttezza (o meno) dei comportamenti di ciascun soggetto possa essere giudicata *ex post* dalla controparte.

Possiamo, infatti, dividere l'*enforcement* endogeno in due principali categorie¹⁴, ciascuno con caratteristiche peculiari; in particolare avremo *enforcement*

1. bilaterale (o interno): le strategie di sanzionamento sono messe in atto direttamente dalle parti coinvolte; il sanzionamento consiste nella cessazione del flusso di quasi rendite derivanti dalla relazione economica.

In questa ottica, le condizioni che garantiscono l'esecuzione del contratto sono:

- la possibilità di osservare le azioni della controparte
- l'esistenza di *surplus* all'interno della relazione
- la ripetizione delle relazione economica (la quale deve avere durata "illimitata" o quantomeno non conosciuta a priori¹⁵)

In questo ambito non è richiesta alcuna trasmissione di informazioni a soggetti terzi.

2. multilaterale (o esterno): questo tipo di *enforcement* è necessario quando la relazione tra le parti ha durata limitata o quando la relazione non è accompagnata dalla creazione di *surplus* (quindi di quasi rendite, pertanto i costi di cambiamento sono nulli).

¹⁴ *Ibid.*, p 393

¹⁵ Se la relazione ha durata nota e l'*enforcing* è bilaterale, la relazione non viene instaurata a priori. Questo aspetto verrà chiarito meglio nei paragrafi successivi.

In questo caso l'*enforcement* è garantito dal comportamento di una terza parte (ad esempio imprese e lavoratori esterni alla relazione economica) che boicotta gli opportunisti, escludendoli da possibili relazioni future: le parti esterne devono poter vedere e interpretare le azioni e devono aver interesse a sanzionare. Il sanzionamento consiste, anche in questo caso, nella cessazione dei flussi di quasi rendite che, però, originano dalla presenza dei soggetti sul mercato e non più dalla relazione economica.

In questo ambito sorge il problema del *free-rider*: il sanzionamento si configura, infatti, come bene pubblico; ciascuno ha interesse che gli opportunisti vengano sanzionati, ma il sanzionare è costoso, quindi nessuno vorrebbe farlo.

In generale, l'*enforcement* endogeno si ottiene con la creazione di istituzioni che prevedono incentivi, per coloro che rispettano i contratti e sanzioni per coloro che, invece, non li rispettano.

La minaccia di sanzionamento non deriva, come ormai chiaro, da un'autorità bensì dal comportamento di una delle due parti (bilaterale) oppure da quello di una terza parte che riceve informazioni in merito e ha interessi economici (multilaterale)¹⁶.

In conclusione, occorre osservare che l'*enforcement* endogeno non può essere implementato nei contratti espliciti (eccessivamente rigidi) ma deve essere basato su contratti dotati di una forte flessibilità e adattabilità, cioè sui contratti impliciti¹⁷.

Una volta chiarito il quadro generale introduttivo, possiamo ora ad analizzare più nel dettaglio i due tipi di *enforcement* endogeno.

¹⁶ *Ibid.*, p 396

¹⁷ Nicita A., Scoppa V., *op. cit.*, p 111

4.1.1 *Enforcement* bilaterale

Consideriamo il caso in cui il sistema di *enforcing* endogeno sia bilaterale; in queste circostanze un soggetto può infliggere direttamente una punizione alla controparte, venuta meno agli accordi contrattuali, senza la necessità che terzi ricevano informazioni in merito. Fondamentalmente, la punizione si realizza tramite l'adozione della seguente strategia:

- interruzione della relazione in caso di opportunismo della controparte;
- rifiuto di intraprendere una nuova relazione con soggetti che si sono mostrati inaffidabili.

Una strategia molto usata è, infatti, quella che prende il nome di “rinnovo contingente” che prevede il licenziamento in caso di *performance* non soddisfacenti: in questo modo viene risolto l'opportunismo dell'agente senza creare quello del principale poiché, in caso di lealtà del lavoratore, il principale non ha motivo di licenziarlo. Inoltre, essendo un accordo bilaterale, non possono configurarsi forme di corruzione¹⁸.

Dal punto di vista di teoria dei giochi, si sta attuando una *trigger strategy* in base alla quale l'accordo viene rispettato fino ad un eventuale opportunismo che implica, in seguito, l'interruzione permanente della relazione economica.

In particolare, la sanzione bilaterale assume forza ed efficacia nel momento in cui valgono certe condizioni, cioè se

1. i soggetti coinvolti possono osservare le azioni intraprese dalla controparte;
2. vi è un orizzonte temporale finito ma incerto, oppure un orizzonte infinito.
(Con orizzonte finito certo, ci sarebbe sempre la spinta a comportarsi in

¹⁸ Scoppa V., *op. cit.*, p 400

modo opportunistico giunti all'ultimo periodo e pertanto, tramite la *backward induction*, si comprende che la relazione economica non verrà neanche avviata). In altre parole, ci deve essere sempre una probabilità positiva che la relazione venga rinnovata nel periodo successivo;

3. esiste *surplus*, quindi quasi rendite positive; deve esistere, cioè, l'incentivo a continuare la relazione (questo avviene sotto varie forme, come ad esempio la promessa del pagamento di un bonus alla fine della vita lavorativa). Le quasi rendite, come precedentemente osservato, sono in contrasto, però, con l'ipotesi di mercati perfettamente concorrenziali.

Un altro problema che, quindi, assume rilevanza all'interno dell'analisi dei contratti impliciti è rappresentato dalle modalità di creazione del *surplus*¹⁹; possiamo individuare alcune tra le più significative fonti, ossia

- la reputazione; se un'impresa ha fama di licenziare con facilità, dovrà proporre elevati salari;
- la disoccupazione di equilibrio e la presenza di posti vacanti (se il lato del mercato dell'agente è razionato, l'interruzione è costosa);
- i costi cambiamento; cambiare partner è costoso;
- gli investimenti specifici necessari per instaurare relazione economica (in particolare il capitale umano).

Definite le “origini” del *surplus*, risulta evidente che è il modo in cui questo viene distribuito che determina la forma del contratto. La “bilateralità” dell'*enforcement* deriva, in sostanza, dalla cessazione dei flussi di quasi rendite generate attraverso le modalità sopra elencate²⁰.

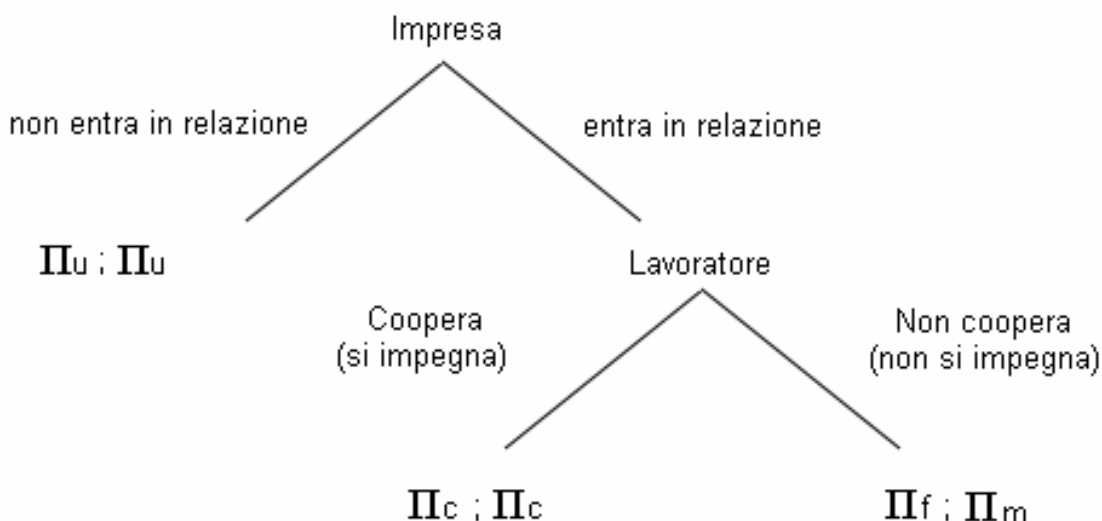
Per comprendere meglio il funzionamento dell'*enforcing* bilaterale possiamo far ricorso allo strumento formale della teoria dei giochi.

¹⁹ *Ibid.*, p 403

²⁰ *Ibid.*, p 404

Rappresentiamo la relazione fra le parti tramite il dilemma del prigioniero unilaterale²¹: il principale (impresa) fa la prima mossa e può entrare o meno in relazione con l'agente (lavoratore).

Da un punto di vista grafico, il gioco può essere rappresentato in questo modo



Se la relazione non viene instaurata, il *pay-off* del principale (impresa) e dell'agente (lavoratore), coincidono con le loro utilità di riserva, che per ipotesi sono entrambe uguali a Π_u .

Analizziamo il caso in cui l'impresa decida, invece, di instaurare una relazione economica; il lavoratore avrà di fronte a sé due scelte, defezionare o cooperare.

Nel primo caso, supponiamo che il lavoratore ottenga Π_m e l'impresa Π_f ; se, invece, il lavoratore coopera, otterrà Π_c , facendo ottenere Π_c anche all'impresa. Ipotizziamo che valga la seguente catena di disuguaglianza $\Pi_m > \Pi_c > \Pi_u > \Pi_f$.

²¹ Questa rappresentazione si distanzia da quella "classica" del dilemma del prigioniero in quanto il gioco è sequenziale e non simultaneo, come invece accade nel gioco standard dal momento che i giocatori scelgono le proprie mosse contemporaneamente. In questo caso, invece, il principale fa la prima mossa e in base a questa, l'agente agirà di conseguenza.

A questo punto, però risulta necessario un chiarimento per non perdere di vista l'oggetto principale d'analisi: i termini "cooperazione" e "defezione" risultano chiari all'interno del gioco, tuttavia sono anche molto generici.

Questa rappresentazione, infatti, può essere applicata a moltissimi ambiti teorici; nello specifico stiamo analizzando la relazione lavorativa, quindi per cooperazione si intende (dal punto di vista del lavoratore) la profusione di un elevato livello di impegno, mentre per defezione, si intende lo *shirking*.

Se il gioco è *one-shot*, oppure ripetuto con orizzonte finito, cioè se il termine è conosciuto, allora tramite la *backward induction*, è facile dimostrare che per il lavoratore è sempre conveniente l'atteggiamento opportunistico; l'impresa, consapevole di questo aspetto, non intraprenderà per nessun periodo la relazione economica, cioè non stipulerà mai un contratto di lavoro con il lavoratore.

Il requisito fondamentale, infatti, è che il gioco abbia un orizzonte infinito o che, pur essendo l'orizzonte finito, sia incerto il termine, ci sia cioè sempre una certa probabilità che il gioco continui.

Al termine di ogni gioco, le parti, per ipotesi, sono in grado di valutare se l'accordo è stato o meno rispettato (in particolare, l'impresa è in grado di verificare il livello di *effort* erogato dai lavoratori) e, quindi, tramite la ripetizione del gioco e l'uso di strategie condizionate, si può arrivare ad un equilibrio cooperativo.

Comportandosi opportunisticamente l'agente ottiene Π_m , risultato che è maggiore rispetto a quello (contingente) che otterrebbe cooperando; bisogna, però, notare che a seguito della defezione, per ogni periodo tranne il primo, otterrà "solo" la sua utilità di riserva Π_u .

Se il lavoratore, invece, si comporta lealmente, otterrà Π_c in ogni periodo. Per capire quale sarà la scelta del lavoratore, quindi, occorre confrontare i *pay-off* attesi.

Dal punto di vista analitico, dato il tasso di interesse r (da cui il fattore di sconto $q = (1 + r)^{-1}$), occorre confrontare il valore attuale dei *pay-off* dell'agente in caso di lealtà, ovvero

$$\Pi_c + q\Pi_c + q^2\Pi_c + q^3\Pi_c + \dots = \Pi_c (1 + r)/r \quad (4.1)$$

con il valore attuale dei *pay-off* in caso di comportamento opportunistico

$$\Pi_m + q\Pi_u + q^2\Pi_u + q^3\Pi_u + \dots = \Pi_m + (\Pi_u/r) \quad (4.2)$$

Il lavoratore, quindi, non si comporterà in modo opportunistico se e solo se vale

$$\Pi_c (1 + r)/r > \Pi_m + (\Pi_u/r) \quad (4.3)$$

da cui si ricava²² facilmente che l'agente è indotto a comportarsi lealmente se vale

$$r < [\Pi_c - \Pi_u] / [\Pi_m - \Pi_c] \quad (4.4)$$

e quindi si ottiene che l'*enforcing* contrattuale endogeno sarà tanto più probabile

- quanto maggiori sono i vantaggi della cooperazione, quanto maggiore, cioè, è Π_c rispetto a Π_u ;
- quanto minori sono i benefici della defezione;
- quanto maggiore è il fattore di sconto (cioè minore il tasso di interesse) e, quindi, quanto maggiore è l' "importanza" (cioè il "peso" attribuito dal lavoratore) dei guadagni futuri.

²² *Ibid.*, p 402

Inoltre, è opportuno sottolineare che, supponendo un adeguato sistema di scambio di informazioni, il principale può essere interpretato anche come una sequenza di soggetti diversi.

Un limite significativo dell'approccio appena esposto, è dato dal fatto che quello cooperativo non è l'unico equilibrio possibile. Gli altri equilibri che si possono avere (non efficienti) sono il risultato delle aspettative errate dei soggetti: il principale, ad esempio, eviterà rapporti con un agente ritenuto (anche erroneamente) inaffidabile; similmente, il principale ottiene un *pay-off* negativo se da fiducia (ingiustificata) ad un agente che non la onorerà²³.

In sostanza, la cooperazione non è una soluzione razionale a meno di non essere in un ambito di ripetizione probabilistica dove i vantaggi della cooperazione e i costi della deviazione sono rilevanti e vi è un fattore di sconto alto.

L'accordo verrà rispettato, quindi, se i benefici (di lungo periodo) della cooperazione eccedono quelli "contingenti" derivanti dall'opportunismo: nell'esempio precedentemente esposto si ipotizza che il guadagno "contingente" della cooperazione sia inferiore di quello, sempre "contingente" dell'opportunismo ($\Pi_m > \Pi_c$) e si trova sotto quali condizioni il contratto viene rispettato; al fine dell'ottenimento di questo risultato l'assunzione fondamentale è rappresentata dall'orizzonte infinito (o finito ma incerto).

Quanto fino ad ora analizzato, però, non si riferisce in modo diretto alla tematica del pagamento di un bonus che l'impresa promette, al fine di garantirsi un elevato livello di impegno da parte dei lavoratori.

Fino a questo momento, il problema principale sembrerebbe riguardare solamente la possibilità da parte del lavoratore, di comportarsi in modo opportunistico; la questione, però, non si riduce solo a questo.

²³ Nicita A., Scoppa V., *op. cit.*, p 122

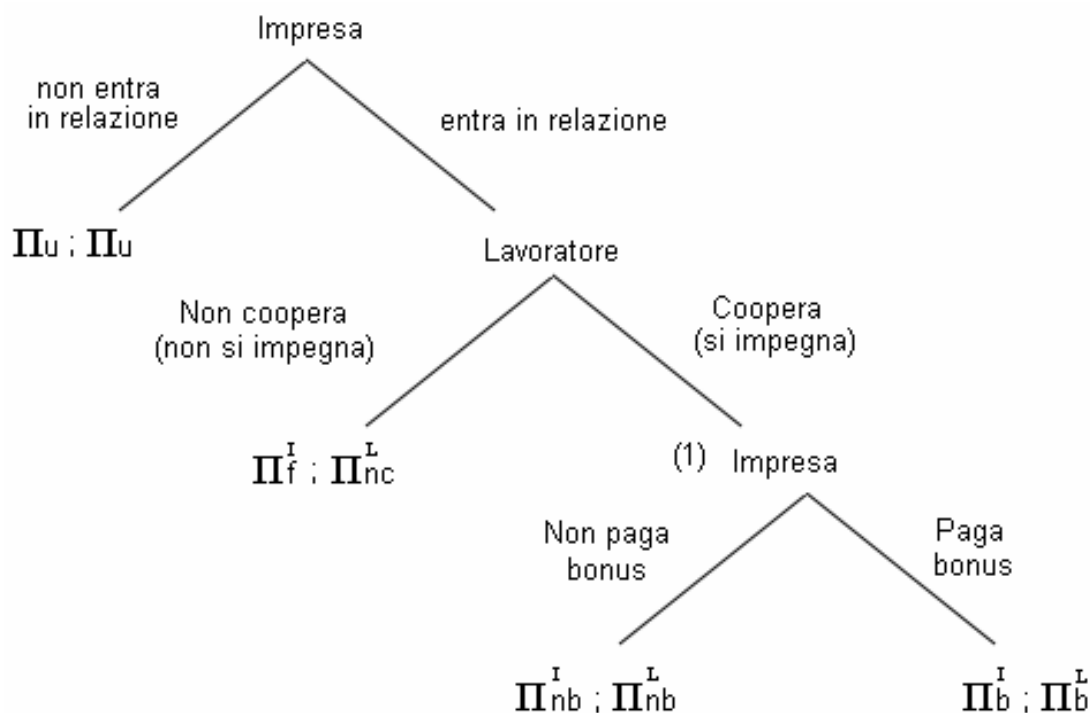
Supponiamo infatti che l'impresa prometta il pagamento di un bonus alla fine di ciascun periodo, come premio per l'elevato *effort* profuso; una volta ottenuto ciò che desidera, l'impresa potrebbe non avere convenienza a rispettare la promessa di pagamento dal momento che il lavoratore, ormai, si è impegnato.

Nei periodi successivi, però, le promesse non verranno più credute; l'impresa, quindi, prima di comportarsi in modo opportunistico, deve valutare con attenzione non solo i costi di breve periodo (pagamento del bonus) ma anche i vantaggi di lungo periodo.

In sostanza, quindi, per scoraggiare il *moral hazard* del lavoratore, l'impresa può adottare una certa strategia, come il pagamento del bonus, che però può indurre al *moral hazard* dell'impresa stessa.

In altre parole, il problema del “doppio” *moral hazard* può essere visto (e risolto) interpretando queste dinamiche come due giochi (analoghi a quello descritto in precedenza) successivi.

Da un punto di vista grafico possiamo impostare il problema in questi termini



Ipotizziamo che nel caso in cui la relazione non venga attivata, entrambe le parti ricevano l'utilità di riserva (per ipotesi uguali a Π_u).

Se la relazione viene attivata, il lavoratore può decidere di impegnarsi oppure di oziare, nel secondo caso il lavoratore godrà di un *payoff* contingente pari a Π_{nc}^L (mentre l'impresa Π_f^I) e successivamente verrà licenziato ricevendo nei periodi successivi sola la propria utilità di riserva.

Supponendo che il lavoratore si impegni, l'impresa potrà decidere se pagare, o meno, il bonus: nel primo caso l'impresa ottiene Π_b^I e il lavoratore Π_b^L ; nel secondo caso, l'impresa e il lavoratore ottengono rispettivamente Π_{nb}^I e Π_{nb}^L .

In particolare, supponiamo che valga la seguente catena di disuguaglianza

$$\Pi_f^I < \Pi_u < \Pi_{nb}^L < \Pi_{nc}^L < \Pi_b^L < \Pi_b^I < \Pi_{nb}^I$$

Se il lavoratore sapesse *ex ante* che la promessa verrà disattesa, in un gioco non ripetuto, gli converrebbe non impegnarsi, ma sapendo questo l'impresa non intraprenderà a priori la relazione.

Se la dimensione temporale è multiperiodale non è detto che il costo di breve periodo (pagamento del bonus) sia maggiore dei benefici di lungo periodo.

Per determinare sotto quali condizioni l'impresa pagherà il bonus occorre andare a studiare (in modo del tutto equivalente all'esempio precedente) i valori scontati dei *payoff* futuri.

Dato il valore attuale del *payoff* in caso di pagamento del bonus²⁴

$$\Pi_b^I + q \Pi_b^I + q^2 \Pi_b^I + \dots = \Pi_b^I (1 + r)/r \quad (4.5)$$

²⁴ Ricordando che $q = (1 + r)^{-1}$

e dato il valore attuale del *payoff* in caso di non pagamento del bonus

$$\Pi_{nb}^I + q\Pi_u + q^2\Pi_u + \dots = \Pi_{nb}^I + \Pi_u/r \quad (4.6)$$

l'impresa, in particolare, avrà convenienza a mantenere le promesse fin quando

$$\begin{aligned} \Pi_b^I (1 + r)/r &\geq \Pi_{nb}^I + \Pi_u/r && \text{cioè} \\ r &\leq (\Pi_b^I - \Pi_u)/(\Pi_{nb}^I - \Pi_b^I) && (4.7) \end{aligned}$$

Supponendo che il lavoratore in un primo momento creda alla promessa di pagamento, sarà incentivato a impegnarsi se il *payoff* in caso di impegno è maggiore (in termini di valori attesi) di quello in caso di shirking, ossia

$$\begin{aligned} \Pi_b^L (1 + r)/r &\geq \Pi_{nc}^L + \Pi_u/r && \text{da cui} \\ r &\leq (\Pi_b^L - \Pi_u)/(\Pi_{nc}^L - \Pi_b^L) && (4.8) \end{aligned}$$

Nel momento in cui risultano verificate congiuntamente la (4.7) e la (4.8), il lavoratore si impegnerà in ciascun periodo consapevole del fatto che all'impresa converrà, di volta in volta, pagare il bonus.

4.1.2 *Enforcement* multilaterale

Come visto nel paragrafo precedente, l'*enforcing* endogeno bilaterale diventa inefficace (cioè le sanzioni dirette possono essere insufficienti) ogni qual volta è consentito all'opportunista di "uscire" senza costo; inoltre, se la relazione ha un orizzonte noto, l'impossibilità di porre in essere un sanzionamento nell'ultimo periodo del rapporto economico, induce a non intraprendere a priori la relazione stessa.

Nelle moderne economie, non sempre si assiste alla creazione di *surplus* all'interno della relazione; i rapporti economici, inoltre, hanno generalmente durata finita e spesso ben nota. Risulta, quindi, necessario che la cooperazione sia garantita da minacce e sanzioni attuate da soggetti esterni alla relazione, attuate cioè da agenti sul mercato (a cui risulta, pertanto, indispensabile l'accesso alle informazioni rilevanti circa il rapporto stesso²⁵).

Spesso, infatti, il *surplus* (quindi i flussi di quasi rendite positive) nasce dalla presenza dei soggetti nel mercato e non dalla relazione economica in se. Quindi il sanzionamento in ogni caso consiste nell'interruzione di un "beneficio", la differenza, però, risiede nella sua origine.

In questo ambito le sanzioni che si possono configurare, hanno una duplice natura, si parla infatti di

- sanzioni di mercato, da parte, cioè, di potenziali partner che si rifiutano di instaurare una relazione con coloro che in passato si sono mostrati inaffidabili e opportunisti;
- sanzioni sociali, l'opportunisto è ostracizzato e quindi è colpito da disutilità diretta. L'*enforcement* sociale (che di fatto si configura come una esternalità) è stato per molto tempo "marginalizzato" nell'analisi economica, tuttavia di recente è stato riscoperto nell'analisi del microcredito dove un insieme di persone è responsabile di tutti gli individui del gruppo.

Le condizioni che consentono l'*enforcing* bilaterale, quindi, si riscontrano con maggiore difficoltà, a causa della frequente mancanza di *surplus* interno alla relazione e/o della durata nota e limitata delle relazioni economiche; in questo

²⁵ *Ibid.*, p 409

contesto, assume un ruolo cruciale (ancor più di prima) il concetto di reputazione, del quale è possibile individuarne due categorie principali²⁶

- (con informazione incompleta) le scelte pregresse trasmettono informazioni circa “l’essere” del soggetto e circa le azioni future;
- (con informazione completa) la reputazione coincide con l’adozione continuata delle strategie d’equilibrio;

la reputazione, quindi, è mantenuta per due scopi: trasmettere informazioni favorevoli e non incorrere in ritorsioni future da parte degli altri giocatori.

In sintesi è possibile definire (nuovamente) la reputazione come quel meccanismo che induce a rinunciare ai benefici di breve periodo in favore di quelli futuri. La reputazione è mantenuta fino a che i benefici futuri (in termini di valore attuale) sono maggiori di quelli contingenti.

L’*enforcement* multilaterale, inoltre, consente di superare alcune delle principali problematiche presenti nei sistemi di *enforcement* esogeno. Un soggetto esterno, avente il compito di valutare il corretto adempimento di un contratto, può non avere le competenze adatte né l’autorità per imporre decisioni, inoltre, a meno di ipotesi ad hoc, non vi è un’attività “d’ufficio” (automatica) ed è necessario il “ricorso” a tale strumento.

Il ricorso ad una corte, inoltre, genera costi legali nonché problemi di corruzione e condizionamento; infine, il sistema legale è eccessivamente “vischioso”, lento negli aggiustamenti, per poter essere adatto a cogliere le dinamiche economiche che spesso sono estremamente rapide.

Affinché l’*enforcement* multilaterale sia efficace è necessario che terze parti siano a conoscenza della relazione economica, conoscano gli accordi, le

²⁶ *Ibid.*, p 410

informazioni rilevanti e abbiano le capacità di interpretare e valutare inadempimenti nonché l'interesse a sanzionare.

Le fasi principali per la realizzazione dell'*enforcement* multilaterale possono essere sintetizzate nei seguenti punti:

- la raccolta e la disseminazione delle informazioni; per avere *enforcement* multilaterale, infatti, è necessario che i costi di trasmissione delle informazioni non siano eccessivamente alti;
- l'analisi e l'interpretazione dei fatti, che può essere suscettibile di errori;
- l'accordo delle aspettative soggettive;
- l'imposizioni delle sanzioni, che comporta il problema del *free-riding*.

A differenza dell'*enforcing* bilaterale, il costo²⁷ del sanzionamento è a carico dei soggetti esterni alla relazione, ma in queste circostanza possono mancare gli incentivi a sanzionare²⁸.

L'attività punitiva ha, infatti, connotati di bene pubblico: tutti traggono beneficio dalla punizione degli opportunisti ma nessuno vorrebbe pagare e quindi si ha il fenomeno del *free-riding*. In alcuni casi, infatti, sono previste sanzioni per chi non sanziona.

Da notare che se l'informazione fosse incompleta, è semplicemente l'operato pregresso che induce ad accordare o meno la fiducia.

Nell'ambito dell'*enforcement* multilaterale molte sono le ipotesi che si possono fare circa l'orizzonte di vita delle parti: ciascuna di queste ipotesi comporta risultati significativamente differenti nell'ambito della teoria dei giochi.

²⁷ Costo che si sostanzia nell'acquisto delle informazioni e nei costi derivanti dal rifiuto di intraprendere altre relazioni nonché nei costi di ricerca.

²⁸ *Ibid.*, p 413

Le ipotesi relative alla dimensione temporale sono fondamentalmente tre²⁹:

- ogni soggetto ha un orizzonte di vita infinito, la relazione economica ha, invece, durata finita a causa del *turnover* tra le parti;
- le relazioni economiche hanno durata finita, i lavoratori hanno un orizzonte di vita finito mentre le imprese hanno vita infinita;
- tutti i soggetti hanno un orizzonte di vita finito.

4.1.2a Relazione con durata finita e orizzonte infinito per entrambe le parti.

Per quanto riguarda la prima ipotesi temporale è opportuno considerare il modello di Milgrom, North e Weingast (1990), utile per studiare il ruolo della disponibilità informativa nell'applicazione dell'*enforcement* endogeno multilaterale³⁰: nel modello viene ipotizzato un numero elevato di soggetti (N) con un orizzonte di vita infinito.

In ogni periodo entrano in rapporto due soggetti alla volta e giocano un dilemma del prigioniero; ciascun soggetto si incontrerà con lo stesso partner ogni N periodi; viene inoltre ipotizzato che i soggetti conoscano solo la propria storia pregressa.

Chi viene ingannato potrebbe attuare la ritorsione ogni N periodi, ma il tasso di sconto necessario per l'*enforcement* bilaterale, se N è grande, tende a zero, quindi a fronte di guadagni immediati, la penalizzazione futura è troppo lontana (pertanto “insignificante”) nel tempo.

²⁹ *Ibid.*, p 414

³⁰ *Ibid.*, p 418

Alla luce di queste considerazioni, si spiega la nascita di istituzioni il cui scopo è quello di diffondere informazioni, coordinare i soggetti e non di sanzionare direttamente.

Sono i privati a stabilire le norme, le sanzioni sono inflitte direttamente dai soggetti che evitano relazioni con partner inadempienti (istituzioni di questo tipo sono le agenzie di rating, gli albi professionali, le agenzie di rilascio licenze). Grazie a queste istituzioni, prende forza un sistema di *enforcement* multilaterale, più facile da ottenere rispetto a quello bilaterale.³¹

E' in questo contesto che sono nate le gilde commerciali, il cui scopo era quello di interpretare i fatti e stabilire quando erano necessarie sanzioni sia contro la città sia contro coloro che non sanzionavano: quest'ultimo sanzionamento è possibile grazie alla creazione di quasi rendite dovute all'affiliazione alla gilda (nell'economia moderna questo approccio è proprio nel mercato del lavoro, ad esempio, dell'ambito sindacale)³².

4.1.2b Relazione con durata finita e orizzonte finito per entrambe le parti.

Tornando alle ipotesi circa l'orizzonte di vita, di particolare interesse è il terzo caso (il secondo in realtà può essere incluso in questo), in quanto, di fatto, rappresenta la situazione maggiormente in linea con la realtà.

I modelli sviluppati nei precedenti contesti temporali, infatti, pur rappresentando un'importa passo avanti nella costruzione di un impianto teorico in linea con la realtà, si basano su ipotesi molto limitative³³.

³¹ *Ibid.*, p 419

³² *Ibid.*, p 420

³³ *Ibid.*, p 427

Nonostante questo, però, la ricerca scientifica è andata avanti ed è riuscita ad ottenere, di fatto, gli stessi contesti temporali partendo, tuttavia, da ipotesi diverse.

L'evidenza empirica, infatti, suggerisce che sia plausibile ipotizzare la coesistenza tra le generazioni di lavoratori anziane e giovani; quindi non esiste un periodo terminale comune.

Questo vuol dire che l'ipotesi secondo cui le imprese hanno un orizzonte di vita infinito, può essere “giustificata” interpretando l'impresa non come un soggetto in sé, ma come un'entità costituita da soggetti che si susseguono nel tempo.

Sulla base di questa constatazione, molti modelli si basano sull'ipotesi che nella fase finale, il lavoratore riceva un “premio” da parte dei giovani che sono, quindi, incentivati a cooperare in vista di questo futuro premio che otterranno a loro volta dalle generazioni successive.

In sostanza, i benefici legati alla fine della vita lavorativa hanno la medesima funzione che aveva il flusso di quasi rendite positive (in un contesto dove l'orizzonte dei soggetti è infinito)³⁴.

In linea con quanto appena detto è utile considerare il contributo³⁵ di Kreps (1990), nel quale viene ipotizzato che le imprese siano gestite da soggetti con vita finita. L'impresa è, di fatto, uno “strumento” che consente di trasmettere la reputazione attraverso le generazioni. La reputazione può essere vista, quindi, come il valore di mercato dell'impresa stessa.

Ipotizziamo che i soggetti vivano due periodi: nel primo acquistano l'impresa e nel secondo la vendono; in entrambi i periodi hanno incentivo a non annullare il

³⁴ *Ibid.*, p 427

³⁵ *Ibid.*, p 427

valore dell'impresa (cioè a non perdere la reputazione) e quindi a non disattendere le promesse fatte ai lavoratori.

L'*enforcement* è, pertanto, garantito sia dai lavoratori sia dai potenziali acquirenti i quali eviteranno rapporti con una impresa che si è mostrata opportunistica e che, quindi, ha perso il proprio "valore".

Come osservato precedentemente, la presenza di relazioni con durata finita e di soggetti con un orizzonte di vita finito, non sono in contrasto con l'idea di un'impresa che ha vita (potenziale) infinita.

Nel modello³⁶ di Cremes (1986) viene ipotizzata un'impresa dove operano T agenti.

Ogni agente vive T periodi ed ogni periodo entra un nuovo agente; in altre parole l'impresa in ogni periodo ha un agente di ogni d'età.

La cooperazione, per il lavoratore, costa e ; inoltre l'attività sinergica di n soggetti produce l'output $Y = n\Pi c$ ³⁷ che verrà diviso tra T soggetti; l' i -esimo otterrà

$$[(n+1) \Pi c]/T - e \quad \text{se coopera} \quad (4.5)$$

$$n\Pi c/T \quad \text{se non coopera} \quad (4.6)$$

Se il gioco è *one-shot*, conviene quasi sempre³⁸ non cooperare; se, invece, il gioco è ripetuto infinitamente si può ottenere cooperazione tranne che per l'ultimo periodo.

Questo risultato si ottiene defezionando nel periodo T e cooperando negli altri, a meno che qualcuno delle generazioni più giovani non abbia defezionato.

³⁶ *Ibid.*, p 428

³⁷ Con $\Pi c > 0$ per ipotesi

³⁸ Tranne nel caso in cui il singolo contributo è "estremamente" produttivo, cioè se $\Pi c > eT$.

Se tutti seguono queste strategia, alla generazione T conviene cooperare; similmente la generazione $T - 1$ otterrà

$$[(T-1) \Pi_c/T - e] + [(T-1) \Pi_c/T] \quad \text{se coopera} \quad (4.7)$$

$$(T-2) \Pi_c/T + 0 \quad \text{se non coopera} \quad (4.8)$$

quindi anche a questa generazione conviene cooperazione poiché, per ipotesi³⁹, si ha $\Pi_c > 0$.

³⁹ *Ibid.*, p 428

4.2 Osservazioni di sintesi sui contratti *self – enforcement*

Come osservato all'inizio di questo capitolo, la teoria dei contratti impliciti assume un ruolo significativo all'interno della letteratura scientifica volta a spiegare ed interpretare le rigidità salariali.

Al di là dei risultati (o dell'impostazione teorica), che possono o meno essere condivisi, un aspetto molto importante, sul quale non si può che concordare, è la rinnovata attenzione verso un mercato, quello del lavoro, che non è più visto come un qualsiasi altro mercato di beni.

Le asimmetrie informative, i costi di transazione, la razionalità limitata degli agenti sono alcune tra le principali cause dell'impossibilità di formulare contratti completi e di ricorrere, quindi, ad una autorità esterna per l'attuazione forzata degli accordi.

L'abbandono dei contratti espliciti comporta, quindi, l'abbandono di un sistema di *enforcing* esogeno che ben aveva “funzionato” in un contesto, però, irrealistico, dove, cioè, era sempre possibile creare un contratto esplicito completo.

Nel nuovo ambito teorico, esiste (ancora) il rischio d'opportunismo da parte dei soggetti coinvolti nella relazione economica, ma è possibile delineare schemi contrattuali con *enforcement* endogeno. Questo “nuovo” tipo di *enforcement* è basato sulle dinamiche tra gli stessi soggetti coinvolti direttamente nella relazione e/o tra soggetti sul mercato (aventi, comunque, un qualche interesse economico riguardo la relazione lavorativa).

E' la continuità della relazione a dar consistenza all'*enforcement* e a consentire di premiare o sanzionare i soggetti coinvolti. L'*enforcement* endogeno dipende:

- dalla disponibilità delle informazioni da parte degli agenti interni e/o esterni alla relazione economica;
- dalla misura e dal genere dei costi che i soggetti sono in grado di imporre alle parti inadempienti;
- dall'interesse di imporre tali sanzioni quando necessarie.

L'*enforcement* bilaterale è realizzato tramite strategie delle controparti: la parte lesa ha interesse a sanzionare quella sleale, per evitare comportamenti opportunistici futuri; l'efficacia dipende dalla continuità e dall'assenza di un termine certo.

Risulta, però, necessaria la creazione di *surplus* all'interno della relazione economica, altrimenti la sanzione avrebbe un costo nullo.

Nell'*enforcement* multilaterale le informazioni devono essere diffuse anche a terzi e le azioni devono essere osservabili e interpretabili. E' necessario un incentivo a sanzionare e a coordinare le aspettative sulle strategie da adottare.

Il maggior pregio dell'*enforcement* multilaterale, consiste nel fatto che mantiene l'efficacia anche se il rapporto ha durata nota e non produce *surplus*; tuttavia, quasi rendite positive sono ugualmente presenti, ma originano dalla permanenza dei soggetti sul mercato⁴⁰.

In particolare, possiamo osservare che l'*enforcement* multilaterale è più diffuso di quello bilaterale qualora

- i costi di diffusione delle informazioni sono bassi;

⁴⁰ L'*enforcement* dipende anche da fattori culturali e sociali, in particolare qualora le sanzioni economiche risultano inefficaci.

- sia agevole il coordinamento tra gli agenti ed esiste interesse a sanzionare;
- esista un'elevata mobilità (*turnover*) del lavoro, cioè quando la durata dei rapporti è breve e i costi di cambiamento sono elevati.

In conclusione possiamo, inoltre, osservare che

- l'*enforcement* endogeno è, per sua natura, distante dal funzionamento dei mercati concorrenziali⁴¹; a causa delle rendite in questi mercati, il salario è diverso da quello di riserva;
- avviare una nuova relazione può essere problematico e richiede, ad esempio, investimenti in informazione; i rapporti saranno, pertanto, di lunga durata ed entrambe le parti preferiranno la stabilità alla mobilità;
- la situazione è, tendenzialmente, non di *market-clearing*: domanda e offerta non si eguagliano ed il meccanismo del prezzo non è in grado di compensare gli squilibri (in termini marginali, infatti, il beneficio non uguaglia il costo); si viene ad avere razionamento del mercato.

Da queste ultime osservazioni è possibile trarre alcune importanti considerazioni circa le dinamiche nel mercato del lavoro: la mobilità del lavoro risulta essere non conveniente né per l'impresa, che deve sostenere dei costi per ricercare e formare un lavoratore adatto a ricoprire il posto vacante, né per il lavoratore.

In particolare quand'anche fossimo in una situazione di piena occupazione, il lavoratore, tendenzialmente, preferirebbe non cambiare lavoro in quanto il rapporto fiduciario è un qualcosa che necessita di molto tempo per potersi creare e mantenere.

⁴¹ A differenza di ciò che accade in concorrenza perfetta, le contrattazioni avvengono solo tra soggetti ben identificati, tra i quali si instaura una relazione "fiduciaria" (intesa in termini economici).

Il problema in caso di disoccupazione è ulteriormente acuito dai lunghi tempi di ricerca di un nuovo posto di lavoro, che possono sfociare nel fenomeno dell'isteresi.

Nel capitolo seguente viene introdotto il sistema dei tornei, basati sul presupposto (realistico) che sia molto difficile, se non impossibile, che un'impresa intrattenga una relazione lavorativa con un singolo dipendente.

In altre parole, i tornei “sfruttano” la pluralità dei soggetti all'interno di una stessa realtà industriale, al fine di costruire uno schema efficiente di incentivazione individuale.

5. *Rank – order Tournaments*: i tornei come schema di incentivazione

Fino a questo momento è stato ipotizzato che i soggetti, coinvolti in modo diretto all'interno di una relazione economica lavorativa¹, fossero soltanto due, ossia, l'impresa e il lavoratore².

Nonostante le forti differenze metodologiche, in tutte le impostazioni analizzate, infatti, il nucleo centrale dell'analisi era la contrapposizione tra impresa e lavoratore, visti come soggetti individuali.

Le imprese, però, contrariamente da quanto sembrerebbe emergere dallo studio di questi schemi, non instaurano mai una relazione economica con un singolo lavoratore³, ma piuttosto con una pluralità di lavoratori che prestano il proprio *effort* fianco a fianco.

In altre parole, un medesimo principale si trova ad avere contemporaneamente una pluralità di relazioni contrattuali con una molteplicità di agenti; l'operato dei singoli lavoratori è solo una componente, un "ingranaggio", che contribuisce al prodotto, quindi alla redditività del principale.

In prima approssimazione, si potrebbe (continuare a) pensare che sia comunque valido rappresentare il rapporto lavorativo come una relazione puramente bilaterale, anche se affiancata a molte altre "identiche": in altre parole, ci sarebbe una pluralità di soggetti "paralleli" e indipendenti l'uno dall'altro.

¹ Anche in alcuni schemi precedenti, l'intervento di soggetti terzi è stato importante, ma solo come soggetti non direttamente coinvolti nel rapporto lavorativo, capaci di garantire l'*enforcement* contrattuale.

² Più precisamente, è stato fino ad ora, ipotizzato che i soggetti a cui il contratto si riferisce fossero solamente due.

³ Questa situazione, di fatto, è impossibile da riscontrare nelle realtà.

L'impegno individuale, però, è "inutile" se non inserito in un quadro complessivo unitario che contempli l'esistenza (o per meglio dire "la coesistenza) di soggetti "complementari" e non semplicemente "paralleli".

I lavoratori si trovano, infatti, a doversi impegnare in modo sinergico per ottenere il prodotto finale: l'*effort* individuale, interagisce con quello degli altri soggetti dando vita (in modo non del tutto prevedibile⁴) al contributo globale al prodotto finale stesso.

A questo punto occorre, però, una precisazione: la sinergia richiesta per poter ottenere il prodotto finale, può essere, infatti, intesa da molti punti di vista.

L'azione dei lavoratori, in particolare, può essere "complementare" in senso stretto o in un senso più generale: nel primo caso, le mansioni di ciascuno, sono differenti (generalmente, in queste circostanza il torneo non viene utilizzato), nel secondo caso, invece, per "sinergia" si fa riferimento ad una situazione generale che si realizza all'interno della stessa impresa, anche (e soprattutto) in presenza di medesime mansioni lavorative. E' a questo ultimo caso che si riferisce il sistema dei tornei.

La redditività del principale, quindi, dipende dallo sforzo "combinato"⁵ di molti agenti che spesso si trovano impiegati in mansioni (*task*) simili e, pertanto, facilmente confrontabili. Questo aspetto può consentire l'individuazione di un sistema d'incentivazione efficiente.

Nei contratti d'agenzia abbiamo visto che la componente stocastica influisce sul prodotto finale e, quindi, sulla remunerazione del lavoratore, il quale viene retribuito sulla base di una qualche valutazione della propria *performance* assoluta.

⁴ E' difficile stabilire quali effetti abbia la combinazione dei vari apporti individuali: questa incertezza può essere vista come una delle componenti stocastiche che incidono sul prodotto.

⁵ Da ora in poi, per "apporto sinergico" si farà riferimento alla seconda accezione, cioè a quella non di complementarità in senso stretto.

In un'ottica più ampia, dove cioè si realizza una cooperazione sinergica⁶, gli agenti operano nello stesso contesto economico e pertanto i fattori casuali sono, in larga misura⁷, comuni e condivisi (questo vale non solo all'interno di una singola impresa, ma può valere anche tra imprese operanti nello stesso settore industriale).

Questo ultimo aspetto è di fondamentale importanza in quanto, all'interno di ciascuna realtà industriale⁸, permette di individuare schemi di incentivazione lavorativa prescindendo (quasi) totalmente da *shock* stocastici comuni.

Determinare la posizione relativa dei soggetti è, in ogni caso, facile e poco costoso; tuttavia, la presenza di *shock* comuni, consente di avere all'interno dello schema retributivo dei tornei (oltre alla riduzione dei costi di rilevamento), anche un maggior livello d'efficienza.

In un contesto di questo tipo, infatti, retribuendo sulla base dei risultati relativi, l'impresa è in grado di filtrare (quantomeno in parte) l'effetto degli *shock* comuni e, tramite la fissazione *ex ante* dei premi, il problema del *trade-off* tra incentivo e ripartizione del rischio risulta notevolmente ridotto.

La dimensione individualista dei lavoratori, pur essendo di fondamentale importanza nell'analisi teorica, non può, pertanto, prescindere completamente dalla dimensione cooperativa e di coesistenza lavorativa, nella quale è indissolubilmente inserita.

⁶ In realtà, l'interazione tra i lavoratori può essere di natura diversa: i soggetti non cooperano necessariamente e possono addirittura intraprendere azioni di sabotaggio. Tale aspetto, però, pur rappresentando un "naturale" sviluppo dei tornei, non verrà approfondito in questo lavoro in quanto "superfluo" ai fini di un'analisi comparativa generale, scopo della tesi.

⁷ L'idiosincrasia degli *shock*, tuttavia, è un aspetto importante dal quale è impossibile prescindere completamente, ma ai fini di un'analisi introduttiva tale problema può essere marginalizzato anche se di fondamentale importanza.

⁸ Il ragionamento può anche essere esteso al di fuori di una singola impresa, creando un torneo che abbia come riferimento posizioni analoghe ma all'interno di altre imprese. Tuttavia, questa è un'ipotesi superflua nello studio introduttivo del torneo.

Questo aspetto ha dato vita, da un lato, a molti filoni di ricerca teorici che hanno spostato l'attenzione dagli incentivi individuali a quelli di gruppo, ma ha consentito anche la creazione di incentivi individuali fondati sull'esistenza di una pluralità di soggetti all'interno della stessa realtà industriale.

I tornei sono nati, appunto, come schema di incentivazione individuale basato, però, sull'esistenza di una pluralità di soggetti in coesistenza, ma anche in concorrenza⁹, tra loro.

Come vedremo più nel dettaglio, i tornei, oltre a filtrare gli *shock* comuni, consentono una notevole riduzione dei costi di transazione (in particolare di misurazione delle *performance*): infatti, è relativamente semplice e poco costoso stabilire quale sia la migliore *performance* relativa.

Passiamo ora ad analizzare più nel dettaglio i principali contributi teorici relativi ai tornei, iniziando con il ripercorrere l'analisi dei due economisti (Lazear e Rosen) che hanno dato vita a questa letteratura.

⁹ Il termine concorrenza, consente di intuire quale sarà uno dei limiti del sistema dei tornei, ossia il pericolo della scarsa operazione tra i lavoratori.

5.1 Pregi e limiti del torneo: il ruolo delle *performance* relative, lo *spread* salariale e il ritorno al *first-best*. Il modello di Lazear – Rosen.

Parlando di tornei come schema di incentivazione lavorativa, risulta evidente il richiamo alle competizioni sportive: ciò che conta è solo la classifica e non il “distacco” tra gli atleti; similmente in ambito lavorativo la *performance* assoluta non ha “rilevanza”.

I primi economisti che si sono addentrati nell’analisi dei tornei come schema di incentivazione, sono stati Lazear e Rosen¹⁰: fino ad allora gli schemi di incentivazione si basavano strettamente sull’individuo in sé e al problema dell’interazione e coesistenza dei lavoratori non veniva assegnato un adeguato peso all’interno dell’analisi teorica.

I vari schemi di incentivazione, dal modello d’agenzia ai contratti impliciti e ai salari d’efficienza, nonostante le enormi e “inconciliabili” differenze d’approccio e d’analisi, si concentrano sull’apporto individuale, in termini assoluti, del soggetto al prodotto o sul livello, anch’esso in termini assoluti, di impegno profuso.

Ponendo, invece, l’attenzione non più sulle *performance* assolute dei singoli lavoratori, bensì su quelle relative, è possibile introdurre un nuovo schema che presenta, come vedremo, notevoli e importanti implicazioni.

Come osservato precedentemente, è sempre facile costruire una classifica relativa, tuttavia è la “condivisione”, tra tutti i lavoratori, del medesimo *shock*¹¹

¹⁰ Lazear, E. P, S. Rosen, *Rank-Order Tournaments as Optimal Labor Contracts*, Journal of Political Economy, pp 841-864. 1981

¹¹ Il problema della componente stocastica torna con forza, ma in realtà assume una rilevanza secondaria, nell’ottica dei tornei, nel momento in cui è comune a tutti i lavoratori; i problemi principali si avranno

stocastico, a fornire a questo schema un grado di efficienza che altrimenti non sarebbe stato possibile ottenere mediante la rilevazione delle *performance* relative.

Lo schema, in sostanza, consiste nel fissare i premi sulla base della posizione in classifica (sulla base, cioè, della prestazione relativa); in questo modo, i lavoratori saranno indotti a impegnarsi maggiormente. Ma qual è l'aspetto che porta ad una maggiore profusione di *effort*?

I tornei, in generale, possono prevedere un premio “contingente” oppure possono riguardare le prospettive di carriera: nel primo caso, il maggior impegno è giustificato dal tentativo di ottenere un premio immediato (come ad esempio una somma monetaria); nel secondo caso, invece, non è importante il premio contingente, quanto le prospettive future.

In questo secondo caso, quindi, il maggior impegno è finalizzato al raggiungimento di una posizione gerarchica superiore (associata, generalmente, ad una remunerazione superiore): il rendimento atteso dei lavoratori che si impegnano di più, si concretizza nella segnalazione delle proprie capacità all'impresa, la quale, *agendo nel proprio interesse*, attua un'operazione d'abbinamento tra le persone che si sono dimostrate maggiormente capaci e le posizioni lavorative più produttive e quindi meglio retribuite.

Il nucleo centrale da sottolineare, però, consiste nel fatto che in entrambi i casi, il lavoratore si impegnerà di più, per massimizzare la probabilità di vittoria.

Un altro aspetto molto importante dei tornei (anche se enfatizzato solo in un secondo momento) riguarda la capacità di questo schema di eliminare il rischio di opportunismo da parte dell'impresa.

qualora così non fosse, in quanto la classifica prestazionale sarebbe sfalzata e non rispecchierebbe le reali capacità individuali.

I premi, infatti, sono fissati a priori in modo chiaro, esplicito ed immutabile, quindi i lavoratori non hanno motivo di dubitare della veridicità delle promesse fatte dall'impresa.

Questo aspetto, però, è emerso solamente in un secondo momento e, come vedremo, permette di focalizzare l'attenzione in particolare sull'ipotesi di neutralità al rischio dei lavoratori.

Risulta evidente, inoltre, che i costi di rilevazione delle *performance* relative sono estremamente più bassi rispetto ai costi di rilevazione delle *performance* assolute.

Nonostante questa maggior facilità, però, se i costi di rilevazione delle *performance* assolute non sono elevati (e se tale operazione è possibile), allora è opportuno inserire nei contratti sia le *performance* relative che quelle assolute ponderando in base a *shock* specifici e *shock* comuni¹².

L'analisi che Lazear e Rosen svolgono nel loro articolo, si sviluppa attraverso alcuni punti chiave:

1. in primo luogo, i due economisti introducono un semplice modello *benchmark*. Nel modello di riferimento, viene ipotizzato che tutti i soggetti siano identici¹³ e neutrali al rischio e viene analizzato il comportamento del lavoratore rappresentativo e dell'impresa, nel caso di remunerazione legata alle *performance* individuali, cioè, quando il contratto prevede un pagamento a cottimo.

I lavoratori ricevono una remunerazione in base a ciò che producono; successivamente, viene delineato (sempre nell'ipotesi di neutralità al

¹² Nicita A., Scoppa V., *op. cit.*, p 83

¹³ I soggetti hanno la medesima produttività, il medesimo salario atteso e profonderanno lo stesso livello di effort ottimale.

rischio dei lavoratori) lo schema di retribuzione basato su una competizione tra lavoratori;

2. la costruzione di un contratto a cottimo, però, può risultare problematica e spesso vengono utilizzate forme contrattuali alternative. Lazear e Rosen, dimostrano la superiorità (in termini di efficienza) del torneo rispetto a contratti alternativi (come ad esempio il salario orario), in quanto viene mostrato che i tornei consentono di giungere ai medesimi risultati del sistema a cottimo senza, però, essere soggetti agli stessi problemi;
3. in ultimo viene dimostrato che sotto l'ipotesi di avversione al rischio dei lavoratori, i tornei possono¹⁴ far raggiungere ad un grado di efficienza superiore rispetto ai salari funzione del prodotto individuale.

La posizione dei lavoratori rispetto al rischio (in particolare il caso di avversione), assume nuovamente notevole rilevanza all'interno dell'analisi teorica.

Analizzando i salari d'efficienza e i contratti impliciti è stato possibile assumere neutralità al rischio dei lavoratori, in quanto, grazie alle caratteristiche strutturali di questi schemi di incentivazione, il problema dell'avversione al rischio non si poneva.

I tornei, però, non sono nati nello stesso ambito degli schemi analizzati nei due capitoli precedenti; l'ipotesi di avversione al rischio, in particolare, è di fondamentale importanza per quanto riguarda alcuni tra i principali risultati del modello di Lazear e Rosen.

Questo modello, quantomeno in un primo momento, si colloca infatti all'interno della "linea evolutiva" del contratto di agenzia: data l'avversione al rischio

¹⁴ Ma questo dipende in grandissima misura dal comportamento delle componenti stocastiche.

dell'agente, né con il contratto d'agenzia né con incentivi lineari è possibile giungere ad una situazione di *first-best*¹⁵; i tornei, invece, possono (anche in caso di avversione al rischio) consentire di giungere ad un livello d'efficienza superiore.

I tornei, però, si sono evoluti anche in un'altra direzione. Successivamente ai primi contributi teorici, infatti, emerse con forza un ulteriore aspetto; grazie alla modalità con cui vengono “costruiti”, infatti, i tornei sono in grado di eliminare l'opportunismo d'impresa: i premi, associati a ciascuna posizione della classifica relativa, vengono fissati ex ante in modo chiaro ed immutabile; in questo modo viene meno la possibilità da parte dell'impresa, di “rimangiarsi” le “proprie promesse”.

Alla luce di quanto appena detto, risulta evidente che l'unico “rischio” a cui si sottopongono i lavoratori, è quello di perdere la competizione; nel “peggiore dei casi” (se, cioè, non vincono la gara), infatti, i soggetti sanno già quanto riceveranno.

Ai fini dell'analisi che intendo svolgere, quindi, è importante chiarire innanzitutto come funziona uno schema incentivante basato su un torneo. A tale scopo è sufficiente analizzare il modello di Lazear e Rosen nell'ipotesi semplificatrice di neutralità al rischio dei lavoratori; passiamo, quindi, a considerarne le principali linee guida.

Nel modello viene ipotizzata¹⁶ la presenza di due lavoratori, ciascuno dei quali contribuisce alla produzione complessiva tramite l'apporto individuale q_i , definito in questi termini

$$q_i = e_i + \varepsilon_i \quad (5.1)$$

¹⁵ In realtà esistono condizioni sotto le quali tale risultato è possibile, tuttavia risultano essere eccessivamente restrittive. Si rimanda al tal proposito al capitolo 1 della tesi.

¹⁶ Lazear, E. P, S. Rosen, *op. cit.*, p 843

dove e indica il livello di *effort* e ε rappresenta una componente stocastica con media nulla, varianza costante e distribuita uniformemente e indipendentemente tra i lavoratori (i.i.d.).

Ipotizziamo che la produzione totale sia data dalla somma delle quantità prodotte dai singoli: a partire da q_i è possibile, quindi, ricavare la funzione di produzione dell'impresa che, nell'ipotesi in cui l'unico fattore produttivo sia il lavoro (pertanto il salario rappresenta l'unico costo sostenuto dall'impresa), è data da

$$Q = Q(q_1 \dots q_n) = q_1 + \dots + q_n \quad (5.2)$$

nel modello, in particolare, viene ipotizzato $n = 2$.

Come in ogni schema di incentivazione, è plausibile ipotizzare che approfondire *effort* sia costoso per il lavoratore; indichiamo con $C(e)$ la funzione di costo del lavoratore, la quale indica per ogni livello di *effort* di quanto diminuisce l'utilità del soggetto. In particolare ipotizziamo che $C(e)$ sia una funzione crescente e convessa nell'*effort*.

Lazear e Rosen considerano in primo luogo il caso di un sistema di remunerazione a cottimo (modello *benchmark*), in cui il salario dipende dalla produzione individuale effettiva¹⁷.

In un primo momento viene ipotizzato, cioè, che i lavoratori siano pagati r per ogni q prodotto; la remunerazione attesa, quindi, al netto del costo dell'*effort*, sarà data dalla seguente relazione

$$E[rq - C(e)] = re - C(e) \quad (5.3)$$

¹⁷ *Ibid.*, p 843

Massimizzando la (5.3) rispetto ad e , si ricava che l'utilità attesa del lavoratore è massima quando vale

$$r = C'(e) \quad (5.4)$$

La massimizzazione per il soggetto, in altri termini, avviene quando al margine il beneficio per il lavoratore (cioè il costo marginale del lavoro dal punto di vista dell'impresa) uguaglia il costo dell'*effort*¹⁸.

Analizziamo ora il comportamento dell'impresa¹⁹: indichiamo con V il ricavo marginale, il ricavo, cioè, per una unità di prodotto; per ogni unità di prodotto le imprese, inoltre, pagano ai lavoratori r .

In termini analitici, i profitti attesi dell'impresa sono dati dalla seguente relazione

$$E(Vq - rq) = (V - r)e \quad (5.5)$$

Nell'ipotesi di concorrenza perfetta sia nel mercato del prodotto che in quello dei fattori, avremo che r tende a V implicando profitti nulli per l'impresa ma anche

$$V = C'(e) \quad (5.6)$$

la situazione che si ottiene in questo caso è quella di *first – best*, infatti il costo marginale dell'*effort* è pari al beneficio (sociale) marginale.

Per quanto riguarda, invece, il torneo, supponiamo che l'impresa al posto di un sistema a cottimo, paghi i lavoratori sulla base delle *performance* relative, elargendo premi fissati a priori.

Il lavoratore maggiormente produttivo²⁰ riceverà un salario $W1$, mentre il “secondo classificato” riceverà un salario $W2$, ovviamente sotto l'ipotesi $W1 > W2$.

¹⁸ *Ibid.*, p 843

¹⁹ *Ibid.*, p 844

Definiamo p appartenente all'intervallo $[0,1]$, la probabilità che un dipendente qualsiasi, che partecipa al torneo, vinca la competizione, con $p(e)$ funzione dell'*effort*.

La funzione di utilità attesa, data la neutralità al rischio, di ciascun lavoratore è espressa in questi termini

$$U = pW1 + (1 - p)W2 - C(e) \quad (5.7)$$

quindi, data la funzione di utilità U dei lavoratori e definita \underline{u} l'utilità di riserva, il vincolo di partecipazione prescrive che deve essere rispettata la disuguaglianza

$$pW1 + (1 - p)W2 - C(e) \geq \underline{u} \quad (5.8)$$

vincolo che, tendenzialmente, sarà verificato come uguaglianza stretta a causa della concorrenza tra i lavoratori. Supponiamo, da ora in poi, che la (5.8) sia sempre verificata.

L'obiettivo dei lavoratori, una volta deciso di accettare il contratto, è massimizzare la funzione di utilità rispetto ad e ; in altri termini, l'obiettivo consiste nell'individuare il livello di *effort* ottimale per i lavoratori.

Derivando la (5.7) rispetto ad e , si ottengono²¹, rispettivamente, la condizione di primo e secondo ordine

$$(W1 - W2) p'(e) - C'(e) = 0 \quad (5.9)$$

$$(W1 - W2) p''(e) - C''(e) < 0 \quad (5.10)$$

²⁰ Determinare quale lavoratore è maggiormente produttivo, rappresenta la fase "operativa" del torneo. Questa fase si concretizza nella redazione, ad opera dell'impresa, della classifica.

²¹ *Ibid.*, p 845

dalle quali si comprende che il livello ottimo di *effort*, che massimizza la funzione di utilità attesa, è funzione dello *spread* salariale; più precisamente l'*effort* ottimale comporta l'uguaglianza tra il costo marginale $C'(e)$ e il beneficio marginale $(W1 - W2)p'(e)$ dello sforzo²².

Di particolare importanza è la relazione tra l'*effort* e la probabilità di vincita del torneo: la relazione tra la funzione di costo del lavoratore e il livello di *effort* è di immediata comprensione, il rapporto tra la probabilità di vincita del torneo e il livello di impegno, invece, necessita di essere adeguatamente chiarito.

Consideriamo due soggetti, i lavoratori j e k ; in particolare, la probabilità p del lavoratore j di vincere il torneo, può essere rappresentata in questi termini²³

$$p = \text{prob} (q_j > q_k) \quad (5.11)$$

cioè, la probabilità che l'output prodotto da j sia maggiore di quello prodotto da k . L'ultima relazione, data la (5.1), può essere così riscritta

$$p = \text{prob} (e_j + \varepsilon_j > e_k + \varepsilon_k) \quad (5.12)$$

che possiamo riscrivere a sua volta, in questi termini

$$p = \text{prob} (e_j - e_k > \varepsilon_k - \varepsilon_j) = G(e_j - e_k) \quad (5.13)$$

Questa ultima relazione afferma che il lavoratore j per poter vincere il torneo deve approfondire una quantità aggiuntiva di *effort*, rispetto al lavoratore k , maggiore del proprio livello di "sfortuna" dovuta alla componente stocastica.

²² Nicita A., Scoppa V., *op. cit.*, p 84

²³ Lazear, E. P, S. Rosen, *op. cit.*, p 845

In altri termini, G è la funzione di distribuzione della variabile casuale $\xi = \varepsilon_k - \varepsilon_j$, pertanto $G' = g (= p')$ è la funzione di densità di ξ , che per ipotesi ha media nulla e varianza doppia rispetto a quella della componente stocastica che influisce sul prodotto individuale²⁴.

Risulta, pertanto, possibile riscrivere la condizione di ottimo del primo ordine in questa maniera²⁵

$$(W1 - W2) g(e_j - e_k) - C'(e) = 0 \quad (5.14)$$

questa ultima relazione rappresenta, in sostanza, la funzione di reazione del soggetto j , il quale determinerà il proprio livello ottimale di *effort* sulla base di quello scelto dal soggetto k (il quale, a sua volta, farà altrettanto in modo del tutto speculare).

Per quanto riguarda le strategie adottate, infatti, viene ipotizzato che ciascun lavoratore determini la propria risposta ottima per ogni livello di *effort* che l'avversario può approfondire (strategia Nash – Cournot²⁶).

In particolare, dal momento che gli individui sono stati ipotizzati identici, risponderanno allo stesso modo allo schema retributivo, da cui si ricava che $e_j^* = e_k^* = \underline{e}$, inoltre in equilibrio, quando cioè $g(0)$ ²⁷, risulta²⁸ che $p = 1/2$.

In equilibrio, infatti, avremo che $(e_j - e_k) = 0$ e il livello d'*effort* è quello che soddisfa la seguente relazione

$$(W1 - W2) g(0) - C'(e) = 0 \quad (5.15)$$

²⁴ In quanto, per ipotesi, entrambe le componenti stocastiche relative ai due lavoratori sono i.i.d. .

²⁵ *Ibid.*, p 845

²⁶ Le dinamiche descritte ricordano, infatti, il modello di oligopolio alla Cournot.

²⁷ $g(0)$ è la probabilità che il livello d'*effort* per i due lavoratori assuma lo stesso valore. In altri termini, $g(0)$ sta ad indicare il caso in cui la probabilità di vittoria dipenda solo dalle componenti stocastiche.

²⁸ *Ibid.*, p 845

il lavoratore, cioè, profonde (come noto) un livello d'*effort* in modo da uguagliare al margine, il beneficio²⁹ ed il costo del maggior impegno.

Essendo i lavoratori identici per ipotesi, risulta evidente che il vincitore è stabilito solo dalla “fortuna”, cioè dagli effetti aleatori; la probabilità *ex ante* di vincita, infatti, è identica e pari a 0,5.

Possiamo inoltre osservare che, data l'ipotesi di convessità della funzione $C(e)$, il costo marginale aumenta, ovviamente, all'aumentare dell'*effort*; quindi, all'aumentare dello *spread*, aumenta il livello di impegno da parte dei lavoratori.

In sostanza, viene individuato il livello ottimo di *effort* (il medesimo per entrambi i lavoratori) che risulta essere funzione dello *spread* tra i due livelli salariali: il livello dei premi³⁰ influenza solo la decisione di entrata e l'utilità attesa del lavoratore in caso di partecipazione.

La condizione necessaria è che, ovviamente, il massimo dell'utilità attesa sia non negativo.

Con il *piece-rate contracts*, abbiamo precedentemente visto che è possibile raggiungere il *first-best*; l'obiettivo del modello di Lazear e Rosen consiste nel dimostrare che tale situazione può essere ottenuto anche attraverso un torneo.

Tale dimostrazione, date le ipotesi semplificatrici di neutralità e identità dei soggetti, risulta particolarmente facile già a partire dalla (5.15): questa relazione indica, definito un certo differenziale salariale, quale è il livello d'*effort* che massimizza l'utilità del lavoratore; nell'ipotesi di neutralità al rischio, l'impresa può fissare liberamente tale differenziale in modo da ottenere il *first-best*.

²⁹ Il beneficio marginale, in particolare, consiste nell'incremento salariale moltiplicato per l'incremento della probabilità di vincere a seguito della profusione di un livello maggiore di *effort*.

³⁰ *Ibid.*, p 846

In sostanza, ricordando che con $V = C'(e)$ si ha la situazione “migliore in assoluto”, è sufficiente porre

$$(W1 - W2) g(0) = V \quad (5.16)$$

per dimostrare che anche con il torneo si può ottenere lo stesso risultato del contratto a cottimo, cioè il *first-best*.

Dalla (5.16), infatti, si comprende che l'impresa, in assenza di “impedimenti”³¹, è in grado di fissare i premi in modo che l'uguaglianza (5.16) sia soddisfatta: conoscendo il tipo di distribuzione di g , infatti, lo *spread* può essere determinato con precisione.

Infatti, fissando il differenziale salariale pari a

$$W1 - W2 = \frac{V}{g(0)} \quad (5.17)$$

si ottiene $V = C'(e)$, ovvero il *first-best*.

In particolare, possiamo constatare la generalità di questo risultato: in una qualsiasi situazione di mercato, data la neutralità al rischio dei lavoratori, il torneo permette, infatti, di giungere alla situazione di *first-best*, senza la necessità che i profitti dell'impresa siano nulli.

Infatti, sono i salari in termini assoluti ($W1$ e $W2$) a dipendere dalle condizioni di mercato, ma ai fini dell'efficienza, è rilevante il loro differenziale che, invece, ne è indipendente: esiste un solo *spread*³² che consente di giungere al *first-best*, ma

³¹ In particolare si fa riferimento all'ipotesi di neutralità al rischio che consente alle imprese di avere pieno potere sulla fissazione dello *spread*. Anche la situazione di mercato, tuttavia, può rappresentare un impedimento, in particolare la *zero profit condition*.

³² Data $g(0)$ ma in particolare, dato il ricavo marginale V .

esistono (potenzialmente) infiniti salari che consentono di ottenere tale differenza.

Se, quindi, non ci sono “restrizioni particolari”, l’impresa può fissare lo *spread* in modo totalmente libero.

Ponendo, invece, la *zero-profits condition*, l’impresa dovrà fissare il differenziale salariale in modo preciso e rispettando determinate condizioni.

In ogni caso (indipendentemente cioè dalla situazione di mercato), la variabile decisionale dell’impresa è rappresentata dallo *spread*, che oltre a garantire la partecipazione deve massimizzare, come precedentemente analizzato, l’utilità attesa.

In base a questi risultati, pertanto, possiamo concludere che, scegliendo adeguatamente i valori di W_1 e W_2 (e quindi il loro differenziale):

- il livello di *effort* che l’impresa ottiene dai lavoratori, tramite un torneo, è uguale a quello che otterrebbe qualora remunerasse le *performance* individuali;
- sia nel contratto a cottimo che nei tornei, la remunerazione attesa è pari al costo dell’*effort*, quindi, l’impresa ottiene il medesimo livello di profitto, che nel caso di concorrenza perfetta è pari a zero.

Essere in concorrenza perfetta, però, non è condizione necessaria per avere l’equivalenza in termini d’efficienza tra il contratto *piece-rate* ed il torneo.

Possiamo, inoltre, operare alcune importanti osservazioni:

- all’aumentare del ricavo marginale, l’impresa attribuirà un maggior valore all’*effort* profuso dai lavoratori; l’impresa, quindi, è indotta ad aumentare lo *spread* salariale. Questo aspetto risulta particolarmente chiaro anche da

un punto di vista analitico, analizzando la (5.17): se V aumenta dovrà aumentare anche lo *spread*.

- all'aumentare della variabilità della componente stocastica che incide sul prodotto, i lavoratori tenderanno a impegnarsi di meno in quanto il prodotto marginale dell'*effort* cala; l'impresa, però, può compensare tale situazione tramite un aumento dello *spread*;
- l'impostazione dei tornei, pur essendo nata nell'ambito dell'analisi teorica che pone particolare importanza all'ipotesi di avversione al rischio, è stata, successivamente, sviluppata anche sotto l'ipotesi di neutralità. In seguito alla nascita di questo filone, infatti, è stato sottolineato che il sistema dei tornei consente di eliminare l'opportunismo in maniera bilaterale: fissando i premi a priori, in modo esplicito ed immutabile, l'opportunismo d'impresa viene evitato così come quello dei lavoratori tramite la competizione.

Lazear e Rosen, in sostanza, arrivano a dimostrare che la remunerazione a cottimo è equivalente ad un sistema basato su un torneo, indipendentemente dalla situazione del mercato cioè dal suo grado di concorrenzialità.

Nonostante questo, però, ritengono che il torneo sia un sistema preferibile qualora le *performance* individuali assolute risultino difficili da valutare; tuttavia, i pregi di questo schema, come vedremo, non lo rendono immune a problemi, alcuni dei quali significativi³³.

Fino ad ora è stato affrontato il problema nel caso di neutralità al rischio; come precedentemente detto l'ipotesi di avversione è centrale nell'analisi di Lazear e Rosen ma, tuttavia, è "trascurabile" ai fini dell'analisi qui svolta.

³³ In un lavoro successivo, Lazear amplia il modello contemplando la possibilità di boicottaggio da parte dei colleghi di lavoro. Alcuni tra i principali difetti del torneo, verranno affrontati nel paragrafo seguente.

Per completezza possiamo elencare alcune osservazioni qualitative riguardanti il caso di avversione al rischio degli agenti (lavoratori)³⁴:

- lo *spread* presenta un “vincolo di dimensione” (come forma implicita di assicurazione per i lavoratori che non accetteranno il contratto se lo *spread* è troppo alto);
- il vincolo dello *spread* può non consentire il raggiungimento del *first-best*; questo dipende, essenzialmente dalla rilevanza degli shock. Se l’incertezza è comune, allora i tornei sono il sistema più efficiente, in caso contrario la remunerazione basata su *performance* assolute è preferibile in modo da evitare, ad esempio, che il salario del soggetto A dipenda, anche, da *shock* propri solo del soggetto B³⁵.

Una volta presentato il modello base nell’ipotesi di neutralità al rischio, per concluderne il quadro, facciamo ora alcune osservazioni di complemento.

Il maggior *effort* che “eventualmente” i lavoratori profonderanno, non determina nell’immediato un aumento di salario e può essere interpretato, entro certi limiti, come un “anticipo” sull’*effort* futuro.

Inoltre, non è detto che i lavoratori siano “attirati” dalla gara: nel modello viene ipotizzato che i concorrenti si impegneranno per vincere solo se il premio³⁶ è sufficientemente grande da compensare il maggior sforzo; d’altronde ciascun lavoratore si impegnerà tanto di più quanto maggiore è lo *spread*, sempre che ritengano di avere una probabilità sufficientemente alta di vincere³⁷.

Il ruolo dello *spread* è, pertanto, fondamentale: data l’ipotesi di neutralità al rischio, lo *spread* può essere fissato a qualunque livello; l’ipotesi di avversione al rischio, invece, comporta, come visto, maggiori problemi.

³⁴ *Ibid.*, p 863

³⁵ Nicita A., Scoppa V., *op. cit.*, p 85

³⁶ Premio che, in realtà, non è il salario maggiore bensì lo *spread* salariale in se.

³⁷ In caso contrario, il lavoratore eviterà a priori di partecipare alla competizione o non accettando il contratto oppure non impegnandosi adeguatamente per vincere.

Nel primo caso, infatti, al lavoratore non interessa se a seguito della sfida, perdendo, guadagnerà poco in termini relativi; nel caso di avversione, invece, la differenza tra le remunerazioni deve essere contenuta in un certo *range*.

L'impresa, ovviamente, cercherà di fissare i premi in modo efficiente e nel far questo dovrà tener conto, qualora sia rilevante, non solo dello *spread* ma anche del *range* massimo.

Un risultato importante del modello è che tanto maggiore è l'incertezza globale sulla produzione (cioè non specifica per un singolo lavoratore) tanto maggiore dovrà essere lo *spread* fissato dall'impresa.

E' vero, infatti, che il torneo consente di "filtrare" la componente stocastica comune, ma è anche vero che all'aumentare di questa, i lavoratori potrebbero essere meno propensi a partecipare alla competizione e pertanto per garantire l'adesione è necessario che l'impresa provveda a correggere il differenziale tramite un aumento dello *spread*.

Questa ultima considerazione si basa strettamente sull'ipotesi di neutralità al rischio, ipotizzando avversione al rischio, invece, tanto maggiore è lo *spread* tanto minore sarà l'incentivo a partecipare³⁸ alla competizione: per i lavoratori, uno *spread* "limitato" rappresenta, infatti, una forma assicurativa implicita.

Se la differenza è troppo elevata, i lavoratori non saranno disposti ad accettare il "rischio" (di perdere) e quindi si asterranno dal competere. Si comprende concettualmente il motivo per cui in caso di avversione al rischio non è detto che si possa ottenere il *first-best*, in quanto l'impresa ha "le mani legate" e non può liberamente fissare il differenziale salariale.

³⁸ Cioè, a non accettare il contratto.

5.2 Osservazioni di sintesi sui tornei e considerazioni aggiuntive

Ricapitolando, da un punto di vista prettamente teorico, un torneo non è altro che una gara, un meccanismo che consente di ordinare i lavoratori in base a regole fissate a priori, in particolare in base alle *performance* relative (più facili e meno costose da rilevare).

L'obiettivo consiste nel costruire uno schema retributivo dove l'elemento fondamentale (sul quale cioè si basano le scelte dei lavoratori e grazie al quale si può giungere al *first-best*) è il differenziale salariale, che può essere ottenuto (potenzialmente) tramite infinite combinazioni di salari che rendono, di fatto, lo *spread* indipendente dalle fluttuazioni economiche.

Le regole (e quindi i premi), fissate a priori, assumono un ruolo centrale nei risultati, consentendo di eliminare il rischio di opportunismo dell'impresa; nelle regole devono essere previsti, infatti, i premi, la modalità di valutazione e i criteri di redazione della classifica.

Fissando questi elementi *ex ante*, quindi, l'unico "rischio" a cui i partecipanti si espongono, è quello di perdere la competizione: i tornei, pertanto, pur essendo nati nell'ambito della letteratura del contratto d'agenzia, permettono di ottenere interessanti risultati anche all'interno dei contratti impliciti, dove gli agenti sono ipotizzati neutrali al rischio.

Sotto tale ipotesi, infatti, l'impresa è totalmente libera di scegliere il differenziale, e pertanto è possibile giungere al *first-best*.

Come osservato in precedenza, il nucleo centrale dell'*enforcing* basato sui tornei è rappresentato dal tentativo di vincere la competizione (cioè dalla

massimizzazione della probabilità di vittoria) tramite la profusione di un maggior livello di impegno; competizione che può prevedere un premio contingente oppure essere relativa alle prospettive future.

La rilevanza del contributo di Lazear e Rosen riguarda fundamentalmente l'introduzione dei tornei come schema efficiente (anche sotto l'ipotesi di avversione al rischio³⁹) che permette di giungere a risultati altrimenti ottenibili anche da altri schemi, consentendo, però, il superamento di alcuni rilevanti problemi.

Nella realtà, però, i concorrenti sono molteplici: l'articolo di Lazear e Rosen non considera questo caso che verrà successivamente ripreso e analizzato da Rosen in studi successivi.

Il problema della molteplicità dei concorrenti è stato oggetto di numerose ricerche successive; il contributo di Jungyoll Yun⁴⁰ ne è un significativo esempio: maggiore è la dimensione della gara, maggiore è anche la diversificazione dei premi.

In particolare, ipotizzando infiniti concorrenti, la posizione in classifica di ciascuno tenderà a “coincidere”⁴¹ con la propria produttività marginale (risultato desiderabile ai fini dell'efficienza): per ogni contratto a cottimo, quindi, esiste un equivalente contratto basato su un torneo, sufficientemente grande, con un numero infinito di concorrenti.

Il contributo di Yun, in particolare, consiste nell'aver dimostrato sotto quali condizioni è possibile ottenere la medesima situazione con un numero finito di

³⁹ In questo caso, non è detto che si possa raggiungere il first-best, comunque si ottiene un livello d'efficienza superiore rispetto ad altri schemi.

⁴⁰ Yun J., *On the Efficiency of the Rank-Order Contract under Moral Hazard and Adverse Selection*, Journal of Labor economics, vol. 15, issue 3, pp 466-494. 1997

⁴¹ Tenderà, cioè, ad essere rappresentativa della produttività marginale. In questo modo ciascuno percepirebbe, similmente al paradigma neoclassico, un salario pari alla propria produttività.

partecipanti, neutrali al rischio, ed un torneo di dimensioni finite; tale risultato si ottiene prevedendo pochi partecipanti ed elevate penalità per chi non si impegna⁴².

Una forma particolare di gara, molto frequente empiricamente, è rappresentata dai tornei sequenziali: una volta finita una sfida, si passa di livello e se ne affronta immediatamente un'altra.

In questo caso l'incentivo, oltre che dall'aumento salariale, è dato anche dal valore atteso dei futuri aumenti di salario (l'incentivo si ha, infatti, anche nel caso in cui nel livello immediatamente successivo non è previsto un aumento salariale⁴³).

I tornei, inoltre, presentano significativi vantaggi rispetto ad altri schemi di incentivazione, rispetto anche a quelli che giungono (come visto nel paragrafo precedente) alla stessa situazione in termini d'efficienza.

In particolare, possiamo elencare i principali pregi di questa tipologia di incentivi; nei tornei, infatti,

- viene evitato il problema dell'opportunismo d'impresa; il premio viene fissato, in maniera immutabile, a priori;
- vengono evitati problemi legati alla riluttanza dei supervisori a dare valutazioni negative;
- il meccanismo di incentivazione è dinamico e si adatta alle condizioni esterne del sistema economico; nell'ambito lavorativo, sia dal punto di vista dell'impresa che del dipendente, la dimensione relativa assume

⁴² *Ibid.*, p 487

⁴³ Questa considerazione può essere estesa fino a "trasformare" il sistema dei tornei in quello dei bonus erogati al termine della vita lavorativa, in questo ultimo caso sorge, però, il problema precedentemente affrontato della credibilità dell'impresa.

un'importante fondamentale e questo fa sì che la variabile chiave sia lo *spread* che può essere adattato, ad esempio, a periodi di recessione;

- la selezione dei lavoratori avviene tramite una valutazione delle capacità relative e questo, oltre a consentire una riduzione dei costi di valutazione, permette un'accurata analisi, in termini relativi.

Il sistema dei tornei, tuttavia, non è esente da difetti e problemi che hanno limitato, nonostante i forti vantaggi in termini di efficienza, la diffusione del sistema stesso; in particolare:

- si può avere il disincentivo alla cooperazione tra colleghi, e vi può essere il pericolo di boicottaggio tra i lavoratori;
- la selezione dei nuovi lavoratori da parte degli *insider*, potrebbe essere distorta in negativo generando un calo del livello medio dei nuovi assunti, al fine di ridurre la “concorrenza” ed evitare che persone valide possano vincere in futuro;
- vi è la possibilità di collusione tra i lavoratori per mantenere un livello prestazionale generale basso così come il rischio di corruzione;
- fino ad ora abbiamo ipotizzato che tutti i lavoratori abbiano le stesse abilità e che l'incentivo all'impegno derivi dal tentativo degli agenti di aumentare la probabilità di vincere.

Se le abilità sono diverse, i lavoratori più abili si dovranno impegnare di meno, mentre quelli meno abili non si impegneranno per niente, ritenendo di non avere possibilità di vittoria.

I tornei, quindi, sono efficaci se e solo se i lavoratori hanno un livello di abilità simile;

- un lavoratore competente ad un certo livello, procedendo nella scala gerarchica potrebbe arrivare ad un livello in cui risulta meno adeguato e abile (principio di Peter).

La scarsa cooperazione, così come il boicottaggio, rappresenta il problema più rilevante dei tornei. Come soluzione sono stati introdotti altri schemi di incentivazione, in particolare gli incentivi nei Team i quali, a loro volta presentano significative problematiche

Se l'incentivo è legato alla *performance* del *team*, la remunerazione media di un membro non terrà conto del proprio grado di abilità: i lavoratori più abili, quindi, tenderanno ad allontanarsi così come (in modo apparentemente strano) anche quelli con abilità inferiori alla media.

Questi ultimi, infatti, pur essendo potenzialmente avvantaggiati dalla situazione, riceveranno un'eccessiva (e ingiustificata) pressione da parte dei colleghi che li spingerà ad allontanarsi.

Un altro importante contributo è, ad esempio, quello di Carmichael (1989)⁴⁴; in questo articolo viene sottolineato che i tornei, nell'ambito dei contratti *self-enforcing*, assumono una connotazione particolare, creando un "terzo soggetto" che produce *surplus* là dove prima non ve ne era⁴⁵.

Supponiamo che nel ciclo della vita lavorativa i dipendenti siano pagati esattamente il salario di riserva e che comportamenti sleali siano prevenuti tramite il pagamento, da parte dei lavoratori stessi, di una certa quantità in caso di "allontanamento" (licenziamento e uscita volontaria).

Un contratto *self-enforcing* è fattibile dato che entrambe le parti guadagnano il loro costo opportunità e il *surplus* necessario per avere l'*enforcement* è creato artificialmente diminuendo la remunerazione dei dipendenti che si comportano opportunisticamente; fino a questo punto, però, niente di nuovo.

⁴⁴ Carmichael H.L., *Self-enforcing Contracts, Shirking, and Life Cycle Incentives*, Journal of Economic Perspectives, pp 65-83. 1989

⁴⁵ *Ibid.*, p 78

Il torneo fa nascere una terza parte che assume, a tutti gli effetti, i connotati di soggetto economico; nasce, quindi, l'ulteriore problema di definirne il comportamento e implementarlo nel contratto *self-enforcing* per evitare sia comportamenti opportunistici sia collusivi.

Il terzo agente è rappresentato, in sostanza, dai lavoratori stessi, uniti in un unico grande gruppo.

Supponiamo che vi siano M lavoratori e che il salario minimo garantito sia W_m , a N lavoratori è garantito un bonus B (che può essere inteso come una promozione⁴⁶), verificabile *ex-post* così come il salario minimo.

La probabilità di essere promossi (cioè di ricevere il bonus) è N/M , mentre il beneficio di comportarsi opportunisticamente è C con probabilità P di essere “sorpresi”,

Il valore atteso dei benefici derivanti da comportamenti disonesti è, quindi, C/P che deve essere inferiore a $(N/M)B$ affinché il comportamento del lavoratore sia in linea con quanto voluto dell'impresa.

Se un lavoratore è opportunist, la “terza parte” si avvantaggerà in quanto sarà più probabile per ciascuno di loro ottenere la promozione.

Carmichael, in sostanza, chiarisce che il *surplus* necessario all'impresa per mantenere la propria reputazione e consentire quindi l'*enforcement* contrattuale, non è necessario nell'ambito dei tornei nel quale le imprese non hanno motivo di essere disoneste⁴⁷.

⁴⁶ *Ibid.*, p 79

⁴⁷ *Ibid.*, p 80

Le imprese che si basano su questo sistema, infatti, riusciranno nel lungo periodo a far uscire fuori dal mercato le imprese che invece devono “necessariamente” produrre *surplus*, competendo con la propria reputazione.

Giunti a questo punto, passiamo all’ultima fase dell’analisi che consiste nel confronto dei diversi schemi di incentivazione a livello macroeconomico.

6. Analisi degli effetti macroeconomici dei sistemi di incentivazione: la disoccupazione involontaria è un “male” necessario per l’enforcement?

A partire dalla crisi petrolifera degli anni settanta, la maggior parte dei paesi sviluppati si allontanò inesorabilmente da una situazione di pieno impiego: l’andamento anticiclico del tasso di disoccupazione, richiedeva l’accurata analisi delle cause della disoccupazione stessa al fine di poter promuovere politiche economiche volte a mitigare il problema.

Sono varie le teorie relative alla disoccupazione e tra queste, molte prendono vita a partire dall’analisi microeconomica del mercato del lavoro: nel corso degli anni si sono succedute diverse impostazioni teoriche che, nonostante le (forti) differenze, ritenevano necessario partire dal livello microeconomico e studiare i comportamenti massimizzanti degli individui a quel livello, prima di procedere all’aggregazione.

La NMK sviluppò approcci finalizzati a microfondare le rigidità salariali in modo da poter spiegare l’esistenza e la persistenza di disoccupazione involontaria; il nuovo paradigma, come noto, era caratterizzato dalla presenza di incertezza e informazione asimmetrica, elementi che possono determinare l’opportunismo dei soggetti coinvolti. Si resero, quindi, necessari strumenti di incentivazione inseriti, in un primo momento, all’interno del “quadro macroeconomico”.

I primi studi riguardanti l’incentivazione lavorativa, infatti, erano volti non tanto alla risoluzione del *moral hazard* in sé, quanto alla spiegazione del perché il salario non scenda, causando la persistenza della disoccupazione involontaria: in questa prospettiva, in realtà, la “risoluzione” dell’opportunismo post contrattuale non era altro che uno “strumento” e non il “fine” dell’analisi.

Quanto appena detto è di fondamentale importanza per chiarire le differenti prospettive da cui ciascun filone teorico si è sviluppato; inizialmente, infatti, gran parte della letteratura scientifica era dominata dall'idea che l'incentivazione dei lavoratori dovesse “produrre” disoccupazione, tramite il pagamento di un salario maggiore di quello d'equilibrio. Questa idea, però, non era frutto di un'analisi “pura”, prettamente microeconomica, riguardante l'incentivazione lavorativa.

Nei capitoli precedenti, infatti, abbiamo posto l'attenzione anche su schemi incentivanti che non implicano necessariamente salari maggiori di quello d'equilibrio e quindi, come vedremo maggiormente in dettaglio, non implicano necessariamente disoccupazione involontaria che, pertanto, non è strettamente necessaria per avere l'*enforcing* contrattuale.

In sostanza, quindi, lo studio delle asimmetrie informative nel mercato del lavoro sembrò, in un primo momento, fornire interessanti strumenti per l'interpretazione di fenomeni economici (la disoccupazione) difficilmente analizzabili all'interno di una singola teoria. Questo aspetto è forte ancora oggi nell'analisi teorica¹, tuttavia, da un certo punto di vista, la sua importanza è stata notevolmente ridimensionata.

Poter ottenere l'*enforcing* contrattuale a prescindere dalla disoccupazione, infatti, sembrerebbe “gettare discredito” sulle teorie volte a microfondare le rigidità salariali: la realtà, però, è estremamente complessa (più di ogni pensabile modello interpretativo); tuttavia, nonostante questo, appare ragionevole muoversi (anche) in altre direzioni nell'analisi della disoccupazione.

In altri termini, la disoccupazione ha molteplici spiegazioni e può essere analizzata da molti punti di vista, ma la prospettiva che più interessa ai fini di questo lavoro riguarda il rapporto tra la disoccupazione e l'*enforcing* contrattuale

¹ Il modello di Shapiro–Stiglitz rappresenta, infatti, una “pietra miliare” nella letteratura riguardante le asimmetrie nel mercato del lavoro.

(analizzato, in una prima fase, nella sola prospettiva microeconomica): la disoccupazione è condizione necessaria affinché l’opportunismo post contrattuale venga risolto? Questa domanda, fondamentalmente, si rende necessaria a causa dell’origine “macroeconomica” di alcuni tra i primi modelli volti ad eliminare il *moral hazard*.

I salari d’efficienza nascono, come ormai chiaro, all’interno della nuova macroeconomia keynesiana e rappresentano uno dei principali strumenti interpretativi delle rigidità salariali da sempre osservate sul mercato del lavoro (sebbene mai pienamente formalizzate e spiegate al livello teorico).

Il modello chiave di riferimento è quello di Shapiro – Stiglitz, modello che, fondamentalmente, implica la presenza e la persistenza di disoccupazione involontaria: l’idea di fondo per eliminare l’opportunismo post contrattuale, consiste nel pagare un salario fisso e nel rendere “costoso” l’eventuale licenziamento facendo in modo che risulti difficile trovare un’occupazione alternativa.

I contratti *self – enforcing*, invece, sono nati come strumento prettamente microeconomico e traggono la propria ragion d’essere dall’impossibilità pratica² di formulare un contratto completo ed esplicito: le parole chiavi, in questo approccio, sono “reputazione”³ e “*enforcement* endogeno”.

La reputazione, ovviamente, acquista un forte significato ed un vero e proprio valore economico solo in un contesto in cui la relazione economica è reiterata. L’*enforcement* endogeno (bilaterale o multilaterale) trova la propria ragion d’essere alla luce dell’incompletezza contrattuale stessa, essendo l’*enforcement* esogeno, il più delle volte, privo di efficacia.

² In realtà, la creazione di un contratto implicito può essere dovuta ad una scelta volontaria delle parti.

³ Oltre alla reputazione, ci sono altri importanti elementi da considerare all’interno di questa impostazione; la reputazione, tuttavia, rappresenta uno degli aspetti fondamentali, se non il principale.

Per ciò che riguarda i tornei (sistema basato sulla premiazione delle *performance* relative), occorre, in particolare, sottolineare che la rilevanza teorica assunta da questo schema è dovuta alla sua diffusione nelle realtà industriali.

Questo schema⁴, infatti, è nato prima nelle imprese e solo in seguito è stato formalizzato a livello teorico: in un certo senso, quindi, potremmo anche affermare che la teoria della “gestione del personale” ha anticipato la “teoria economica” dandogli un’importante spunto di riflessione.

Dopo aver visto quale è la “dimensione” microeconomica di questi schemi, il successivo passo dell’analisi, obiettivo ultimo di questo lavoro, consiste nell’andare a vedere quale di questi sistemi di incentivazione sia maggiormente “efficiente”⁵ a livello macroeconomico.

Nello studio dei contratti d’efficienza è stata posta una domanda molto importante che è opportuno riproporre nuovamente dal momento che non ha trovato (ancora) una risposta completa e definitiva; la domanda è “ come è possibile che scelte razionali a livello microeconomico, si traducano in scelte (apparentemente) irrazionali a livello macroeconomico?”.

La domanda sembra del tutto lecita dal momento che la disoccupazione, di fatto, è “creata” volutamente essendo, in tal contesto, un effetto di una scelta razionale: in realtà, però, tentare di dare una risposta a questa domanda può risultare sbagliato.

In primo luogo, la realtà, pur dovendo necessariamente essere studiata tramite modelli, non è mai “sintetizzabile” in una singola rappresentazione, inoltre bisogna andare a vedere l’origine ed il motivo d’essere delle varie teorie.

⁴ Originario, anch’esso della dimensione puramente microeconomica.

⁵ Quale, cioè, implica il minor livello di disoccupazione involontaria pur facendo ottenere i medesimi risultati (ovvero l’impegno lavorativo) a livello microeconomico.

Senza nulla togliere al modello di Shapiro – Stiglitz, occorre sottolineare che, in realtà, questo è un modello costruito ad hoc per spiegare un fenomeno (la disoccupazione) che, quindi, “non poteva altro che non essere il risultato del modello stesso”.

I contratti *self – enforcing*, ma ancor di più i tornei, nascono, invece a livello puramente microeconomico ed è forse per questo che è stato possibile superare l'apparente paradosso concretizzato nel fatto che la razionalità individuale sembra implicare l'irrazionalità “globale”.

In conclusione, forse, la domanda precedentemente fatta, non “merita” risposta semplicemente perché cambiando la prospettiva d'analisi (partendo, cioè, da una prospettiva puramente microeconomica) in realtà, viene privata del proprio significato.

L'ultimo passo dell'analisi consiste nello studio degli effetti macroeconomici (più precisamente “occupazionali”) dei vari schemi di incentivazione, in modo da compararne l'efficienza sociale e stabilire, finalmente, se la disoccupazione è necessaria per l'*enforcing*.

Per svolgere questo confronto, nel corso della trattazione verranno ripresentate le impostazioni precedenti sotto ipotesi più stringenti in modo da semplificare l'analisi.

Per quanto riguarda i salari d'efficienza, verrà ripreso il modello di Shapiro – Stiglitz in una veste ulteriormente semplificata. I contratti impliciti ed i tornei verranno formalizzati utilizzando le medesime notazioni e strumenti (*asset value equation*⁶) usate nei salari d'efficienza, in modo da poterne confrontare i risultati in modo chiaro e diretto.

⁶ Questo strumento è già stato utilizzato in precedenza, nell'analisi del modello di Shapiro – Stiglitz: per una spiegazione più dettagliata si rimanda al capitolo 3 pagina 81 della tesi.

6.1 Confronto dei diversi schemi di incentivazione: l'analisi tramite le *asset value equation*

Giunti a questo punto, il passo successivo da compiere (obiettivo ultimo della tesi) consiste nell'analizzare e confrontare gli effetti macroeconomici dei vari schemi di incentivazione.

Una volta terminato lo studio a livello microeconomico, sarebbe stato possibile, già da allora, operare alcune importanti osservazioni circa gli effetti sulla disoccupazione di ciascuno di essi (cosa che per altro, è stata parzialmente fatta in modo, però, non sistematico).

Per dare rigore formale all'analisi, tuttavia, è opportuno modellare i diversi schemi attraverso la medesima notazione, al fine di poter ottenere una rappresentazione omogenea, basata sulle stesse convenzioni, e quindi confrontabile in modo diretto ed esplicito.

Al fine di poter ottenere questa situazione, consideriamo il lavoro di Meccheri⁷, nel quale vengono esposti, sintetizzati e confrontati i diversi schemi di incentivazione sia dal punto di vista microeconomico sia dal punto di vista degli effetti macroeconomici.

Nell'articolo vengono, in primo luogo, definite le assunzioni di base relative al contesto economico, alle condizioni informative e ai flussi di disoccupazione⁸ in un equilibrio di *steady-state*.

Supponiamo che vi sia un numero molto elevato di imprese (identiche tra loro) e di lavoratori (omogenei); ipotizziamo, in particolare, che la forza lavoro (L) sia

⁷ Meccheri N., *Employment with Alternative Incentive Schemes when Effort is Not Verifiable*, Labour, vol. 19(1), pp 55-80. 2003

⁸ Alcune di queste ipotesi sono già state fatte e precedentemente analizzate.

sufficiente a soddisfare qualsiasi livello della domanda di lavoro da parte delle imprese. La relazione tra l'impresa e il lavoratore viene rappresentata, formalmente, tramite un gioco ripetuto in tempo discreto ($t = 1, 2, \dots$), gioco che verrà reiterato fino alla cessazione del rapporto.

Ipotizziamo, per semplicità, che entrambe le parti possano godere di un periodo di vita infinito, che utilizzino lo stesso tasso di interesse r e che siano neutrali al rischio. La funzione di utilità del lavoratore, data l'ipotesi di neutralità al rischio, può essere rappresentata semplicemente come il salario al netto dell'*effort*

$$v = w - e \quad (6.1)$$

Per quanto riguarda l'*effort* profuso, invece, ipotizziamo che i lavoratori possano scegliere tra due diversi livelli di impegno, e_h e e_L tali per cui $e_h > e_L$; per ipotesi l'utilità di riserva è data da $\underline{v} > 0$.

Dal punto di vista dell'impresa, invece, la funzione di produzione è ipotizzata alla Solow⁹, cioè del tipo

$$Y = f(en) \quad (6.2)$$

dove n indica il numero di lavoratori impegnati in una generica impresa mentre e indica l'*effort*; in particolare viene ipotizzato che la funzione di produzione abbia $f' > 0$ e $f'' < 0$ (produttività marginale decrescente).

Supponiamo, inoltre, che se il lavoratore scegliesse il livello e_L allora il livello di produzione sarebbe nullo; il livello di impegno più basso, in altri termini, è totalmente improduttivo¹⁰.

⁹ Nel modello di Solow, il parametro d'efficienza e non è identificato con il livello di *effort*, tuttavia, ipotizzare che l'efficienza si possa ottenere con il maggior impegno, rappresenta un modo per microfondare il concetto di efficienza stessa.

¹⁰ Successivamente, per *shirking* si intenderà specificatamente la profusione del livello di impegno e_L .

Data la (6.2), in particolare, $f'(e_h n)e_h$ rappresenta la produttività marginale del lavoro quando tutti gli n dipendenti, approfondono il livello di impegno più alto.

Dal punto di vista sociale, inoltre, è sempre conveniente assumere un numero di lavoratori strettamente positivo; virtualmente, infatti potrebbe esistere un livello di occupazione tale per cui la produttività marginale risulti minore del costo opportunità. Al fine di evitare questa situazione ipotizziamo che, per n che tende a zero (al tendere, cioè, al livello massimo della produttività marginale), valga

$$f'(e_h n)e_h - e_h > \underline{y} = 0 \quad (6.3)$$

La funzione dei profitti dell'impresa, invece, è data da

$$\Pi = Y - \sum w_i \quad \text{con } i=1..n \quad (6.4)$$

dove l'unico costo di produzione è rappresentato dal monte salari. Nel modello viene, infine, ipotizzato che la relazione lavorativa termini (in quanto divenuta improduttiva) con probabilità esogena s .

Per quanto riguarda la dimensione informativa¹¹, ipotizziamo che *l'effort* (unica misura della *performance* lavorativa) sia osservabile da parte dell'impresa che è sempre in grado di valutare la correttezza dei comportamenti dei lavoratori; l'osservabilità, però, non si estende ad una autorità esterna che, quindi, non può intervenire; si rendono, pertanto, necessari strumenti di *enforcement* endogeni.

Quali sono le principali caratteristiche distintive dei tre differenti schemi?

- Nei contratti d'efficienza, il salario è indipendente dalle prestazioni e l'*enforcement* è garantito dal costo del licenziamento;

¹¹ *Ibid.*, p 58

- Nei contratti impliciti (che prevedono un bonus) assume una rilevanza fondamentale il concetto di reputazione (le informazioni relative alla reputazione devono, quindi, essere trasferibili e distribuibili a tutte e tra tutte le parti);
- Nei tornei quel che conta è la prestazione relativa e la possibilità di vincita aumenta lo sforzo profuso.

Ipotizziamo, in particolare, che la reputazione dei lavoratori non sia trasferibile nel mercato¹²; gli agenti esterni, quindi, sono ignari se il licenziamento deriva da motivi esogeni o endogeni¹³.

Anche i lavoratori licenziati, quindi, hanno sempre una probabilità positiva di trovare un altro impiego. I lavoratori, inoltre, sono in grado di valutare se le “promesse” (pagamento del bonus) dell’impresa sono state mantenute. Questo avviene, in particolare, grazie al fatto che i lavoratori più anziani trasmettono a quelli più giovani le informazioni rilevanti circa “l’onestà” dell’impresa; il medesimo effetto si avrebbe in presenza di un’organizzazione sindacale.

Per quanto riguarda il flusso della disoccupazione, ipotizziamo che N sia il numero totale di occupati in un certo periodo: in equilibrio stazionario il flusso in entrata e in uscita nel mercato del lavoro deve essere uguale¹⁴.

I lavoratori che diventano disoccupati sono sN , dove s indica la probabilità (esogena) di diventare disoccupato; i disoccupati che trovano lavoro, invece, sono $m(L - N)$ dove m è la probabilità di trovare lavoro, da cui in equilibrio si avrà $m(L - N) = sN$ e quindi $m = sN / (L - N)$ e sapendo che $u = (L - N)/L$, possiamo riscrivere la probabilità m di trovare un lavoro in questi termini

$$m = s(1 - u) / u \quad (6.5)$$

¹² *Ibid.*, p 59

¹³ Gli agenti esterni, cioè, non sanno se il soggetto in questione è stato licenziato o si è allontanato volutamente.

¹⁴ Questa analisi è già stata fatta nel capitolo 3, ma è stata reinserita per completezza e poiché in questo capitolo verranno fatte alcune ipotesi semplificatrici rispetto ai modelli già analizzati.

Sfruttando le *asset value equation*, possiamo dunque determinare il valore scontato atteso delle vita lavorativa di un soggetto disoccupato

$$\begin{aligned} rV_u &= \underline{v} + m(V_e - V_u) \quad \text{da cui} \\ V_u &= [\underline{v} + mV_e] / (r + m) \end{aligned} \quad (6.6)$$

dove V_e indica l'utilità attesa scontata di un lavoratore impiegato in un tempo generico; in sostanza la prima uguaglianza dice che rV_u deve essere uguale all'utilità di riserva sommata a ciò che il lavoratore otterrebbe dall'attività di ricerca di un nuovo lavoro¹⁵.

Se la forza lavoro (L) è tale da soddisfare ogni livello di domanda di lavoro, allora la probabilità dell'impresa di trovare un nuovo lavoratore è prossima a all'unità: in questo contesto è del tutto plausibile assumere che l'intero potere di mercato sia dell'impresa che può fissare il salario minimo compatibile con il più alto livello di *effort*¹⁶.

Prima di procedere nella formulazione dei diversi schemi di incentivazione, è opportuno costruire un modello *benchmark* di riferimento. Supponiamo di essere in una situazione in cui si ha *enforcement* esogeno: i lavoratori verranno remunerati esattamente in base all'*effort* profuso e, in particolare, profonderanno il livello e_h fin quando il salario è *maggiore o uguale* all'utilità di riserva sommata alla disutilità dovuta all'elevato *effort* ($w \geq \underline{v} + e_h$).

La massimizzazione dei profitti, quindi, comporta $w = \underline{v} + e_h$. L'impresa, pertanto, impiegherà un numero di lavoratori tale da uguagliare w e la produttività marginale del lavoro

$$f'(e_h)n e_h = \underline{v} + e_h = w \quad (6.7)$$

¹⁵ *Ibid.*, p 60

¹⁶ *Ibid.*, p 61

e risulta di immediata comprensione che la (6.7) descrive la situazione “migliore in assoluto” (*first-best*).

Fatto il quadro introduttivo, passiamo a confrontare i tre principali schemi di incentivazione: l’analisi verrà fatta sotto ipotesi maggiormente restrittive rispetto a quelle precedentemente introdotte; questo consentirà di semplificare l’analisi (e di non essere, quindi, forviati da un’eccessiva “pesantezza analitica”), senza nulla togliere alla generalità dei risultati.

6.1.1 *Shirking model, implicit contracts* (bonus) e tornei: analisi semplificata sfruttando le *asset value equation*

Le idee alla base dei tre principali schemi di incentivazione sono state sufficientemente chiarite e approfondite nei precedenti capitoli, quindi l’obiettivo di questo paragrafo consiste solo nel trovare uno strumento che consenta di rappresentare i tre schemi in modo “omogeneo” e pertanto confrontabile.

Come visto in precedenza, i diversi approcci, pur volti alla risoluzione del medesimo problema di fondo, presentano significative differenze sul piano strutturale a causa, in particolar modo, dei differenti ambiti teorici in cui sono nati. Queste differenze, pur non impedendo un confronto qualitativo, non consentono di operare un confronto formale.

Un modo per rendere uniforme l’analisi e poter, quindi, confrontare (non solo concettualmente) i risultati, consiste nell’utilizzo delle *asset value equation*, strumento già introdotto nell’analisi del modello di Shapiro–Stiglitz.

L'analisi consiste nel determinare, per ognuno dei diversi schemi di incentivazione, i valori V_h e V_L , rispettivamente il valore scontato atteso nell'arco della vita lavorativa in caso di elevato impegno e il valore scontato atteso nell'arco della vita lavorativa in caso di scarso impegno (*shirking*).

In questo modo, quindi, è possibile ottenere risultati analitici coerenti ed uniformi tra loro.

6.1.1a Salari d'efficienza

Partiamo con il considerare il modello di Shapiro–Stiglitz¹⁷ in una veste semplificata rispetto alla formulazione presentata nel capitolo 3; in particolare, rispetto al modello originario, si assume che uno *shirker* venga sempre “scoperto” durante atteggiamenti opportunistici e per questo licenziato.

Analiticamente il problema viene così impostato

$$\begin{aligned} rV_h &= \hat{w} - e_h + s(V_u - V_h) \quad \text{da cui} \\ V_h &= [\hat{w} - e_h + sV_u] / (r + s) \quad (\text{impegno}) \end{aligned} \quad (6.8)$$

$$\begin{aligned} rV_L &= \hat{w} - e_L + (V_u - V_L) \quad \text{da cui} \\ V_L &= [\hat{w} - e_L + V_u] / (1 + r) \quad (\text{shirking}) \end{aligned} \quad (6.9)$$

Il lavoratore, quindi, si comporterà in modo corretto se $V_h \geq V_L$, inoltre, ricordando la (6.6), si ricava (condizione di compatibilità agli incentivi)

$$\hat{w} \geq \underline{y} + e_L + (e_h - e_L) [(1 + r + m) / (1 - s)] \quad (6.10)$$

¹⁷ *Ibid.*, p 61

le imprese, quindi, pagheranno il salario minimo compatibile con questa disuguaglianza¹⁸ che in equilibrio verrà verificata come uguaglianza stretta; dalla (6.10) risulta chiaro che, in questo caso, $\hat{w} > \underline{y} + e_h$, inoltre, sostituendo \hat{w} nella (6.8) e nella (6.9), in equilibrio risulterà che

$$V_h = V_L > V_u \quad (6.11)$$

da cui si comprende che i lavoratori preferiscono strettamente lavorare al salario \hat{w} piuttosto che essere disoccupati; la disoccupazione in equilibrio risulta, quindi, essere involontaria.

L'impresa, pertanto, impiegherà un numero di lavoratori tale da uguagliare la produttività marginale e \hat{w} , in termini analitici

$$f'(e_h n) e_h = \underline{y} + e_L + (e_h - e_L) [(1 + r + m) / (1 - s)] \quad (6.12)$$

dove m è stato precedentemente endogeneizzato (6.5). Si ottengono, pertanto, i risultati tipici del modello di Shapiro – Stiglitz.

6.1.1b Contratti *self – enforcing*

Per quanto riguarda i contratti impliciti, ipotizziamo che vi siano due tipi di accordo¹⁹: il primo contratto (esplicito) stabilisce che i lavoratori ricevano un salario indipendente dall'*effort*, il secondo contratto (implicito), invece, stabilisce che i lavoratori ricevano, in caso di elevato impegno, un bonus.

Nel primo caso l'*enforcement* contrattuale è esogeno mentre nel secondo è di natura endogena.

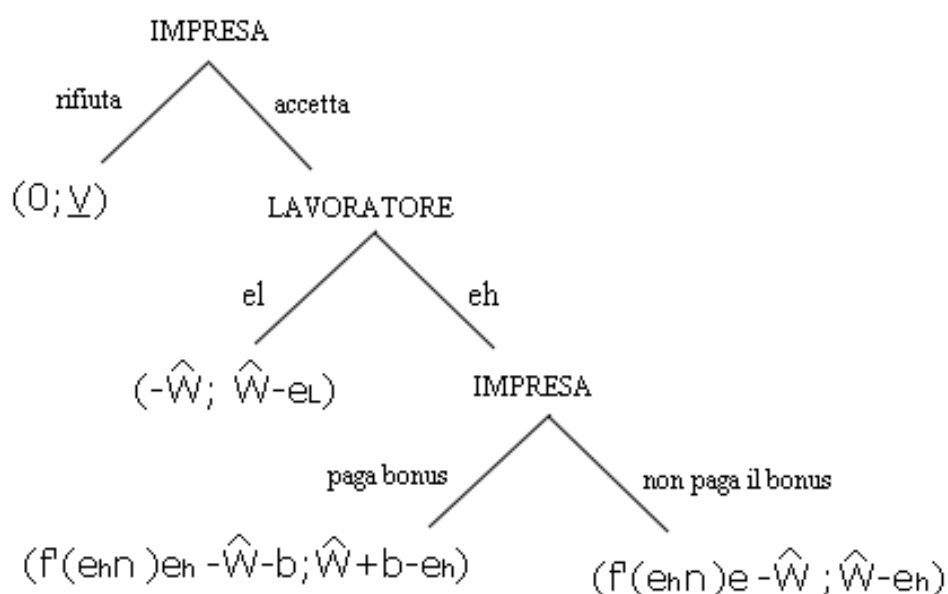
¹⁸ La (6.10), in particolare, rappresenta una “variante” della NSC del modello originale.

¹⁹ *Ibid.*, p 62

Nel caso in cui il bonus non venga corrisposto, i lavoratori, nei periodi successivi,profonderanno il livello più basso di *effort* e la produzione, per ipotesi, sarà nulla.

Il potere di mercato, inoltre, è tutto in mano all'impresa che, quindi, può fissare il salario ed il bonus al livello minimo possibile²⁰.

Sfruttando la rappresentazione ad albero della teoria dei giochi, la relazione tra le parti (in un periodo) può essere così formalizzata²¹



Consideriamo, in primo luogo, il caso *one-shot*: il potere di mercato è interamente in mano all'impresa la quale fisserà il salario al livello minimo tale da compensare la disutilità per il basso *effort* e l'utilità di riserva ($w = \underline{v} + e_l$).

In questo caso, infatti, il salario non svolge il ruolo “incentivante” (svolto, invece, dal bonus), il suo scopo è quello di far accettare il contratto; è per questo motivo che il salario deve compensare (oltre che all'utilità di riserva) e_l e non e_h .

²⁰ *Ibid.*, p 63

²¹ Questa rappresentazione è del tutto analoga a quella presentata nel capitolo 4 p 122; in questo caso, però, i *payoff* sono esplicitati in modo più dettagliato.

Il *payoff* del lavoratore in caso di elevato *effort* è dato da $w + b - e_h$, in caso di basso *effort* $w - e_L$, per cui il bonus sarà $b = e_h - e_L$ ²².

Se il gioco è *one-shot*, in particolare, l'impresa si comporterà sempre in modo opportunistico non pagando il bonus; il lavoratore, consapevole di questo fatto, approfondirà il livello più basso di *effort*, ma questo induce l'impresa a non instaurare a priori la relazione stessa. Questo risultato appare evidente sfruttando semplicemente la *backward induction*.

Se invece il gioco viene ripetuto, le parti devono considerare l'eventuale punizione derivante da comportamenti sleali: se l'impresa è sleale, allora il lavoratore erogherà e_L oppure interromperà la relazione; se è il lavoratore ad essere sleale, l'impresa negherà il bonus o licenzierà il dipendente.

In altre parole, l'impresa dovrà soppesare con attenzione i costi di breve periodo (pagamento del bonus) e i vantaggi di lungo periodo nel caso in cui il costo di breve periodo venga sostenuto (vantaggio che consiste nell'incremento della produttività dovuto al consolidamento della reputazione).

Ovviamente, nel caso in cui il lavoratore profonda al massimo e_L , all'impresa conviene sempre terminare il gioco, per motivi d'efficienza²³.

Similmente a prima, possiamo trovare i valori V_h e V_L nel modo seguente:

$$\begin{aligned} rV_h^{IC} &= \hat{w} + b - e_h + s(V_u - V_h^{IC}) && \text{da cui} \\ V_h^{IC} &= [\hat{w} + b - e_h + sV_u] / (r + s) && (\text{impegno}) \end{aligned} \quad (6.13)$$

$$\begin{aligned} rV_L^{IC} &= \hat{w} - e_L + (V_u - V_L^{IC}) && \text{da cui} \\ V_L^{IC} &= [\hat{w} - e_L + V_u] / (1 + r) && (\text{shirking}) \end{aligned} \quad (6.14)$$

²² *Ibid.*, p 63

²³ *Ibid.*, p 64

quindi, non si avranno comportamenti opportunistici da parte del lavoratore se $V_h^{IC} \geq V_L^{IC}$; ricordando, inoltre, la (6.6), la compatibilità agli incentivi comporta che

$$b \geq e_h - e_L \quad (6.15)$$

per non avere comportamenti sleali, sembrerebbe quindi sufficiente un bonus tale da coprire la disutilità del maggior *effort*.

Le imprese, ovviamente, pagheranno il bonus minimo consistente con la (6.15) che verrà, quindi, soddisfatta come uguaglianza; nella situazione di equilibrio, pertanto, i lavoratori riceveranno esattamente il costo opportunità, cioè $\hat{w} + b = \underline{y} + e_h$ ²⁴ da cui si ricava che

$$V_h = V_L = V_u \quad (6.16)$$

questo tipologia di contratto, quindi, implica un livello nullo di disoccupazione involontaria²⁵.

Come sappiamo, però, nel caso di contratti impliciti sorge il problema della credibilità d'impresa, base dell'*enforcement* contrattuale endogeno²⁶. Risulta, quindi, necessario andare a studiare le dinamiche decisionali all'interno della realtà industriale per giungere ad un risultato conclusivo; in altre parole, l'impresa deve essere credibile nel promettere un certo livello di bonus.

Come visto in precedenza, la credibilità non si può avere in un gioco *one-shot*, in quanto il costo di breve è sempre maggiore del beneficio di lungo periodo (che di fatto è nullo, in un contesto uniperiodale): la reputazione, come noto, può essere costruita solamente in un contesto di ripetizione.

²⁴ Formalmente avremo $\hat{w} = \underline{y} + e_L$ e $b = e_h - e_L$, da cui $\hat{w} + b = \underline{y} + e_h$

²⁵ *Ibid.*, p 65

²⁶ Il problema dell'inconsistenza intertemporale della promessa di pagamento del bonus è già stato affrontato, dal punto di vista, teorico, nel capitolo 4 p 121.

Se il gioco è ripetuto, dobbiamo, pertanto, andare a vedere quando è conveniente per l'impresa, rispettare le promesse; per far questo, occorre analizzare i profitti dell'impresa in ciascuna situazione (sfruttando anche in questo caso le *asset value equations*²⁷), analiticamente avremo

$$\begin{aligned} r\Pi_{nc}^{IC} &= f'(e_h n)e_h - \hat{w} - b && \text{da cui} \\ \Pi_{nc}^{IC} &= [f'(e_h n)e_h - \hat{w} - b] / r && \text{(bonus pagato)} \end{aligned} \quad (6.17)$$

$$\begin{aligned} r\Pi_c^{IC} &= f'(e_h n)e_h - \hat{w} - \Pi_c^{IC} && \text{da cui} \\ \Pi_c^{IC} &= [f'(e_h n)e_h - \hat{w}] / (1 + r) && \text{(bonus non pagato)} \end{aligned} \quad (6.18)$$

e quindi l'impresa pagherà il bonus se $\Pi_{nc}^{IC} \geq \Pi_c^{IC}$, cioè se

$$f'(e_h n)e_h \geq \hat{w} + b(1 + r) \quad (6.19).$$

Considerando congiuntamente sia la condizione affinché i lavoratori si impegnino e quella affinché le imprese paghino il bonus, si ottiene la condizione generale (necessaria) di incentivo-compatibilità²⁸

$$f'(e_h n)e_h \geq \underline{v} + e_L + (e_h - e_L)(1 + r) \quad (6.20)$$

Nel contesto dei contratti impliciti, quindi, pagando $\hat{w} + b = \underline{v} + e_h$ si ottiene lo stesso risultato (in termini occupazionali) del modello *benchmark*; tuttavia, fino a che la (6.19) non è verificata, le promesse dell'impresa risultano non credibili e pertanto i lavoratori scelgono di erogare il livello più basso di *effort* (improduttivo per l'impresa).

Dal punto di vista dell'impresa occorre, infatti, considerare sia il “costo” di breve periodo (bonus) sia il “vantaggio” di lungo periodo (l'incremento della

²⁷ *Ibid.*, p 65

²⁸ *Ibid.*, p 66

produttività): al livello occupazionale del modello *benchmark* la produttività marginale del lavoro, nel caso dei contratti impliciti, risulterebbe troppo bassa perché il vantaggio di lungo periodo possa essere maggiore del costo di breve periodo (quindi possa essere credibile la promessa di pagamento del bonus).

In altre parole, il livello di *first-best* non è “sostenibile” in termini reputazionali e occorre pertanto determinare il livello *minimo* della produttività tale per cui il costo di breve periodo risulti inferiore del vantaggio di lungo: questo livello della produttività è definito dalla (6.20), verificata come uguaglianza.

In conclusione, quindi, l’impresa paga $\hat{w} + b = \underline{y} + e_h$ ma il livello d’occupazione che le “impedisce” di essere opportunistica è inferiore a quello di *first-best*; il meglio che l’impresa può fare è fissare un livello d’occupazione che soddisfi la (6.20) come uguaglianza.

Data la (6.19), infatti, risalta immediatamente all’attenzione il termine $(1 + r)$: se l’impresa non avesse problemi di credibilità, potrebbe fissare l’occupazione in modo da soddisfare semplicemente la relazione $f'(e_h n) e_h = \hat{w} + b = \underline{y} + e_h$, giungendo al *first-best*.

Il termine²⁹ $(1 + r)$, però, accresce il valore destro della (6.19) e, quindi, la produttività marginale deve uguagliare un valore maggiore rispetto a $(\hat{w} + b)$: essendo la produttività marginale decrescente³⁰ risulta, quindi, evidente che l’occupazione deve essere inferiore rispetto al caso *benchmark*, pur essendo corrisposto lo stesso ammontare.

²⁹ Dovuta al contesto di ripetizione.

³⁰ In questo contesto, di fondamentale importanza, è l’ipotesi di produttività marginale del lavoro decrescente.

6.1.1c *Rank–Order Tournaments*

Per quanto riguarda il torneo, ipotizziamo che il vincitore riceva W , mentre il secondo classificato riceva w (dove, ovviamente $W > w$); ipotizziamo, inoltre, che chi profonde un livello basso di *effort* non vinca mai³¹ e venga licenziato alla fine del periodo, chi invece profonde il livello alto, ha una probabilità p positiva di vincere.

Ipotizziamo per semplicità che $w = \underline{y} + e_L$; anche in questo caso, quindi, il salario w non svolge un ruolo incentivante (svolto invece dal torneo), inoltre, a differenza dei contratti impliciti, di per sé non ha neppure lo scopo di garantire l'adesione al contratto, funzione assolta, invece, dal differenziale salariale.

Seguendo lo stesso schema di ragionamento svolto in precedenza, avremo³²

$$\begin{aligned} rV_h^T &= pW + (1 - p)w - e_h + s(V_u - V_h^T) && \text{da cui} \\ V_h^T &= [pW + (1 - p)w - e_h + sV_u] / (r + s) && \text{(impegno)} \end{aligned} \quad (6.21)$$

$$\begin{aligned} rV_L^T &= w - e_L + (V_u - V_L^T) && \text{da cui} \\ V_L^T &= [w - e_L + V_u] / (1 + r) && \text{(scarso impegno)} \end{aligned} \quad (6.22)$$

Pertanto, nell'ambito del torneo, i lavoratori profonderanno un elevato livello di impegno se $V_h^T \geq V_L^T$ da cui, ricordando la (6.6), otteniamo (condizione di compatibilità agli incentivi)

$$W \geq \underline{y} + e_L + (e_h - e_L)(1/p) \quad (6.23)$$

³¹ Questa ipotesi, in particolare, è di fondamentale importanza per i risultati che si ottengono e deriva da quella in base alla quale l'impresa conosce perfettamente l'impegno lavorativo dei dipendenti.

³² *Ibid.*, p 67

quindi, per motivare i lavoratori, è “sufficiente” un premio $(W - \underline{w} - e_L)$ elevato: in equilibrio la relazione (6.23) è verificata come uguaglianza e il salario atteso³³ per i lavoratori può essere riscritto in questi termini

$$pW + (1 - p)w = \underline{w} + e_h \quad (6.24)$$

In equilibrio inoltre avremo

$$V_h = V_L = V_u \quad (6.25)$$

quindi, anche in questo caso, la disoccupazione è solo volontaria.

A differenza dei contratti impliciti, però, l’impresa non può comportarsi in modo opportunistico perché i premi sono fissati a priori; in altri termini, l’impresa non ha problemi di credibilità per quanto concerne le sue promesse sui pagamenti salariali.

Ciascuna impresa assumerà poi un numero di lavoratori tale da uguagliare la produttività marginale del lavoro e il salario, in questo caso, atteso, per cui:

$$f'(e_h n) e_h = pW + (1 - p)w = \underline{w} + e_h \quad (6.26)$$

Si noti come, dal punto di vista occupazionale, l’espressione (6.26) comporti la medesima situazione del modello *benchmark* (*first – best*)

³³ Dalla (6.23) si ricava, infatti, $W - w = (e_h - e_L)p^{-1}$ da cui $pW = pw + (e_h - e_L)$ e quindi aggiungendo e sottraendo $w = \underline{w} + e_L$ e operando alcuni semplici passaggi algebrici, si ottiene la (6.24).

6.1.2 Analisi degli effetti macroeconomici

Passiamo, a questo punto, ad operare il confronto vero e proprio: per ipotesi le imprese sono identiche, quindi prenderanno le stesse decisioni di salario e di occupazione, in particolare, tutte le imprese hanno lo stesso sistema di incentivazione.

Se i vincoli di compatibilità agli incentivi sono soddisfatti, il lavoro è allocato in modo efficiente tra le imprese³⁴.

La funzione di produzione totale è rappresentata da

$$Y = F(e_h N) = \max \sum f_j(e_h n_j) \\ \text{con } \sum(e_h n_j) = e_h N$$

la funzione di produzione aggregata, implica una domanda aggregata di lavoro con le stesse proprietà di quella singola.

I salari rappresentano “semplicemente” un trasferimento di surplus tra agenti; in equilibrio il *surplus* totale sarà dunque pari a

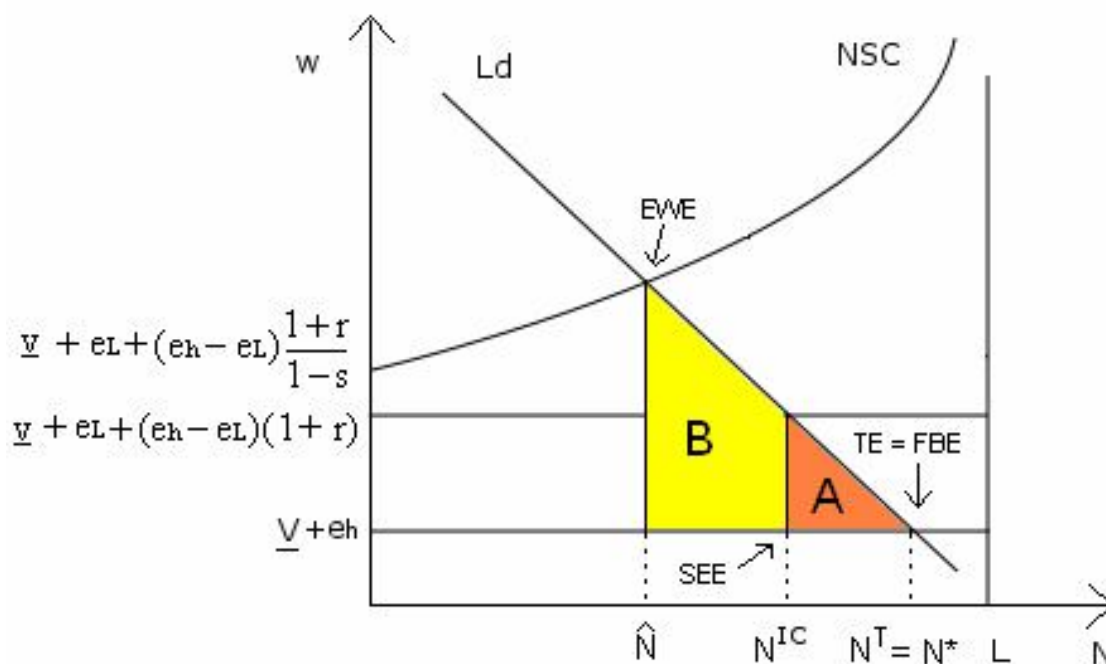
$$S = F(e_h N) + N(-e_h) + (L - N)\underline{v} \quad (6.27)$$

Massimizzando S rispetto N , la condizione del primo ordine sarà data da

$$F'(e_h N^*) e_h = \underline{v} + e_h \quad (6.28)$$

dove n^* è il livello di occupazione di *first-best* che si ottiene se tutte le imprese pagassero $\underline{v} + e_h$ e impiegassero n lavoratori in modo che la produttività marginale del lavoro uguagli $\underline{v} + e_h$; questa situazione è rappresentata graficamente dal punto FBE

³⁴ *Ibid.*, p 68



Nel caso dei salari d'efficienza, in base alla NSC, il salario pagato da una singola impresa è dato da

$$\hat{w} = \underline{y} + e_L + (e_h - e_L) \{ [1 + r + (s(1 - u) u^{-1})] / (1 - s) \} \quad (6.29)$$

Attuando l'aggregazione tra tutte le imprese, si ottiene

$$\hat{f}'(e_h N) e_h = \underline{y} + e_L + (e_h - e_L) \{ [1 + r + s(1 - u) u^{-1}] / (1 - s) \} \quad (6.30)$$

situazione rappresentata dal punto EWE, individuato (come noto) dall'intersezione tra la NSC e la curva di domanda di lavoro aggregata $(F'(e_h N))^{35}$.

Nel caso dei contratti impliciti, invece, le condizione di incentivo-compatibilità (per le imprese e per i lavoratori), stabiliscono che

$$\hat{f}'(e_h n) e_h = \underline{y} + e_L + (e_h - e_L)(1 + r) \quad (6.31)$$

³⁵ *Ibid.*, p 69

quindi, una volta operata l'aggregazione, avremo un livello di occupazione pari a N^{IC} in quanto complessivamente verranno impiegati lavoratori in base a

$$f'(e_h N^{IC}) e_h = \underline{y} + e_L + (e_h - e_L)(1 + r) \quad (6.32)$$

situazione che graficamente è rappresentata dal punto SEE, dal momento che viene corrisposto ugualmente l'ammontare $\underline{y} + e_h$.

Le imprese avrebbero convenienza a raggiungere N^* (ottenendo in più l'area A come *surplus*) ma la promessa del bonus non sarebbe credibile; in altre parole, per poter essere credute, le imprese devono rinunciare ad una parte dei “guadagni” contingenti³⁶ che dovranno essere minori o uguali rispetto al valore scontato atteso di quelli futuri.

In sostanza, vi è una perdita d'efficienza dovuta al mancato sfruttamento dei guadagni a causa di un sistema di *enforcement* imperfetto: la perdita, infatti, è tanto maggiore quanto maggiore è il tasso di interesse (inversamente proporzionale all'occupazione d'equilibrio). A questo punto possiamo già osservare che i contratti impliciti sono “più efficienti” dei salari d'efficienza, nonostante implicino essi stessi un certo livello di disoccupazione involontaria.

Per quanto riguarda invece i tornei, il salario atteso è $pW + (1 - p)w = \underline{y} + e_h$, ottenuto a partire dalla condizione di compatibilità agli incentivi (6.23); a livello di singola impresa, il numero di soggetti che viene assunto è tale da soddisfare la relazione

$$f'(e_h n) e_h = \underline{y} + e_h \quad (6.33)$$

che implica un livello di occupazione aggregato pari a N^T , dato da

³⁶ In questo contesto sarebbe più corretto dire “accettare maggiori costi contingenti in vista di benefici futuri”.

$$f'(e_h N^T) e_h = \underline{v} + e_h \quad (6.34)$$

situazione equivalente a quella ottenuta in presenza di *enforcement* esogeno; i tornei, quindi, consentono di ottenere nel mercato del lavoro, una situazione di *first-best*: rispetto ai contratti impliciti si “guadagna” (in termini di efficienza) l’area A, mentre rispetto ai salari d’efficienza l’aria A+B.

In conclusione occorre ribadire due principali aspetti: i risultati ottenuti si basano strettamente sull’ipotesi di rendimenti decrescenti del lavoro e sull’ipotesi di perfetta osservabilità dell’*effort* da parte dell’impresa.

Questa ultima ipotesi, in particolare, consente di ottenere il *first-best* nel caso dei tornei; in presenza di ipotesi maggiormente realistiche (quando, cioè, l’*effort* non è perfettamente osservabile dall’impresa) anche i tornei non consentono di raggiungere la situazione “migliore in assoluto” e generano anch’essi disoccupazione involontaria.

In particolare, come mostrato nelle estensioni del lavoro di Meccheri³⁷, si può venire, ugualmente, ad avere disoccupazione involontaria, anche a seguito dell’implementazione del torneo, poiché, in presenza di osservabilità imperfetta dell’*effort*, uno *shirker* (nonostante il suo atteggiamento opportunistico) ha comunque una probabilità positiva di vincere la competizione (possibilità che precedentemente veniva espressamente esclusa).

Tuttavia, anche in un contesto di questo tipo, il risultato rilevante consiste nel fatto che la “classifica d’efficienza” dei diversi schemi di incentivazione, resta invariata³⁸: il torneo resta, in ogni caso, il sistema d’incentivazione relativamente più efficiente³⁹.

³⁷ *Ibid.*, p 71

³⁸ Occorre tuttavia notare che nell’analisi fino ad ora svolta, non sono stati considerati i limiti e i difetti del torneo (cap 4); considerando anche tali aspetti non è detto che si ottengano i medesimi risultati.

³⁹ *Ibid.*, p 73

7. Sintesi della tesi e commenti conclusivi

Senza ombra di dubbio, la teoria dei contratti ha dato vita ad uno dei filoni di ricerca più importanti e promettenti nell'ambito economico.

Fino alla scuola neoclassica (inclusa), nessuno aveva pensato che l'informazione fosse un vero e proprio bene economico, con un suo mercato e, quindi, con un suo "prezzo".

L'idea che tutta l'informazione fosse perfettamente e gratuitamente disponibile, ha prodotto tutta una serie di importanti risultati teorici alcuni dei quali, però, sono rimasti "tali", cioè astratti concetti teorici.

Ben presto, infatti, venne abbandonato l'impianto neoclassico e si cominciò a trattare l'informazione come un qualcosa non più distribuito in modo uniforme; questa evoluzione, consentì di "modellare" un fenomeno molto frequente in economia, ossia l'opportunismo dei soggetti (differente dal "semplice" egoismo).

L'opportunismo delle parti può concretizzarsi, in ambito economico e in presenza di asimmetrie informative, attraverso la selezione avversa o tramite il *moral hazard*; in entrambi i casi si ha "fallimento" del mercato, cioè l'allontanamento da una situazione efficiente e si può anche arrivare alla scomparsa del mercato stesso (Akerlof, 1970).

Come evitare l'opportunismo? Come impedire la potenziale fine dei mercati? Come giungere ad una situazione grazie alla quale tutti i soggetti coinvolti sono in grado di trarre il massimo vantaggio possibile date le circostanze?

Nell'ambito della teoria dei contratti, la risposta a queste domande è una sola, "tramite un contratto adeguatamente costruito".

Queste sono alcune tra le domande cardine dell'economia dei contratti e sono le domande alle quali ho cercato di dare una risposta, facendo riferimento, in particolare, all'ambito lavorativo e passando in rassegna la letteratura riguardante tale settore.

Per i soggetti, lavorare è “necessario” ma “disutile”: ciascun soggetto, entrando nel mondo del lavoro definisce, in modo (praticamente) automatico e a seguito di considerazioni morali, un livello minimo di impegno che intende approfondire. Questo livello, però, può non essere uguale a quello desiderato dall'impresa; quest'ultimo, in particolare è, tendenzialmente, superiore¹. In assenza di incentivi adeguati, il lavoratore si limiterebbe a approfondire il livello di impegno da lui ritenuto il “minimo necessario”, dal momento che tale sarebbe la scelta economicamente più razionale. Come esser sicuri che i lavoratori assecondino le richieste di impegno dell'impresa?²

¹ Tendenzialmente il livello di impegno ritenuto “moralmente accettabile” è strettamente positivo.

² Riflettendoci bene, infatti, anche nel modello neoclassico, i lavoratori non hanno nessun motivo per impegnarsi poiché potrebbero (senza costi e in modo immediato) trovare un'occupazione alternativa di periodo in periodo.

7.1 Sintesi dei diversi schemi di incentivazione

7.1.1 Contratto d'agenzia: riepilogo

Uno dei primi approcci analizzati è il modello d'agenzia, all'interno del quale la principale problematica è rappresentata dal *trade-off* tra incentivo e ripartizione dei rischio.

La redditività del principale (datore di lavoro) dipende dall'azione dell'agente (lavoratore); quest'ultimo, in particolare, può approfondire diversi livelli di impegno ed è nell'interesse del principale creare un contratto che induca l'agente a approfondire esattamente il livello desiderato dal datore di lavoro.

Il lavoratore, in particolare, cerca stabilità e certezza salariale (per ipotesi è avverso al rischio) e quindi, tendenzialmente, preferirà un salario fisso piuttosto che uno aleatorio³. Il datore di lavoro, invece, può diversificare il portafoglio contrattuale, per cui è lecito ipotizzarlo neutrale al rischio.

Il risultato principale del modello d'agenzia, consiste nel legare la remunerazione al prodotto finale in modo da garantire l'impegno del lavoratore: in questo modo (sino al caso limite di *residual claimant*, dove l'agente è titolare del reddito residuale al netto del *pay-off* del principale) il lavoratore, sicuramente, profonderà il livello di impegno richiesto dall'impresa.

Il prodotto finale, però, oltre che dall'impegno del lavoratore, dipende anche da variabili casuali (aleatorie), per cui la remunerazione del lavoratore diventa anch'essa aleatoria.

³ Più precisamente, date due alternative (la prima produce un risultato certo, la seconda produce in *media* lo stesso risultato della prima), un soggetto avverso al rischio preferirà l'alternativa certa.

Questo aspetto determina l'inefficiente allocazione del rischio, dal momento che, per ipotesi, il lavoratore ne è avverso. Viceversa, puntando solo all'efficienza allocativa (ponendo, cioè, tutto il rischio a carico dell'impresa), il lavoratore (retribuito con un salario fisso) non avrà motivo per impegnarsi, profonderà, cioè, il livello di impegno da lui ritenuto "moralmente" accettabile.

Nella pratica, quindi, ciò che si realizza è una situazione di *second-best*, caratterizzata da inefficienza allocativa e produttiva, dal momento (che a causa dell'avversione al rischio) gli incentivi sono attenuati. Questo approccio, tuttavia, serve solo a chiarire la problematica microeconomica di fondo e non ha avuto particolare seguito a livello macroeconomico.

I contratti d'agenzia, inoltre, prevedono (di fatto) la possibilità di ricorrere ad una corte per far valere in modo forzoso le clausole contrattuali: il problema fondamentale di questa impostazione è rappresentato "solo" dal *trade-off* tra incentivo e ripartizione del rischio. Inoltre, questa impostazione si basa in larga parte sull'utilizzo di indicatori delle *performance*; negli approcci successivi, invece, si cercherà di formulare un contratto direttamente sull'*effort* (e questo diventa possibile, in quanto cambia la prospettiva d'analisi).

Negli schemi di incentivazioni successivi, infatti, il problema principale riguarda la risoluzione del *moral hazard* in un contesto dove una corte non ha autorità (dove, in particolare, non esistono indicatori oggettivi delle performance) e dove, quindi, è necessario far in modo che il lavoratore si impegni senza per questo provocare "tentazioni" opportunistiche dal lato dell'impresa.

Nei salari d'efficienza, il *moral hazard* del lavoratore è risolto senza creare quello dell'impresa, tramite il pagamento di un salario fisso. Nei contratti *self-enforcing*, invece, il problema del doppio *moral hazard* torna in rilievo e viene risolto in un contesto multiperiodale. Anche nei tornei, l'opportunismo post-contrattuale dei lavoratori è risolto senza creare quello dell'impresa.

7.1.2 Salari d'efficienza: riepilogo

I salari d'efficienza sono nati all'interno della NMK ed hanno come scopo la microfondazione delle rigidità salariali, al fine di spiegare il fenomeno macroeconomico della presenza e della persistenza della disoccupazione involontaria (“impossibile” da giustificare in puro ambito neoclassico).

In assenza di indicatori oggettivi delle *performance*, non è possibile ricorrere ad una corte per assicurare l'*enforcement* contrattuale; è in questo ambito che opera la teoria dei salari d'efficienza, dove l'*enforcement* è garantito dalla “minaccia” della perdita del flusso di quasi rendite associato alla condizione di occupato.

Nel contesto neoclassico, un aumento salariale (a parità di tutte le condizioni di mercato) rappresenta solo e soltanto un aumento dei costi di produzione.

All'interno della teoria dei salari d'efficienza, invece, si ritiene che un aumento dei salari (nonostante determini, in ogni caso, un aumento dei costi) possa indurre ad un aumento dell'efficienza produttiva, identificata in un secondo momento nell'impegno lavorativo. Si viene, quindi, ad identificare il *trade-off* tra il pagamento di un elevato salario e l'ottenimento di un livello maggiore d'efficienza.

All'interno di questo filone, uno dei modelli principali di riferimento è quello di Shapiro–Stiglitz (1984), un modello intertemporale continuo di azzardo morale: tramite un sistema di monitoraggio imperfetto, vi sarà una probabilità positiva di essere sorpresi durante un comportamento opportunistico.

Considerando congiuntamente la probabilità endogena di trovare un lavoro e quella esogena di perderlo, nonché il valore scontato atteso della vita lavorativa nei diversi casi in cui un lavoratore si può trovare (disoccupato, occupato che si impegna, occupato che non si impegna), è possibile determinare il livello

salariale minimo (in modo da massimizzare i profitti) tale per cui si avrà l'impegno lavorativo (grazie al fatto che tale livello è maggiore rispetto al costo-opportunità del lavoratore).

Questo livello salariale, quindi, implica un certo livello di disoccupazione involontaria, che, pertanto, è utilizzata come strumento "punitivo" in quanto aumenta il costo che ciascun lavoratore è costretto a sostenere in caso di licenziamento per motivi disciplinari: i lavoratori saranno leali perché in caso di licenziamento difficilmente troveranno immediatamente un'occupazione alternativa allo stesso salario (di efficienza).

Sviluppi più recenti, però, si sono soffermati (anche se con modalità profondamente differenti) sull'incentivazione puramente microeconomica, non dando peso, almeno inizialmente, agli effetti macroeconomici. In ogni caso, questi sviluppi hanno mostrato come alcuni schemi nati in un'ottica prettamente microeconomica possano produrre importanti risultati in termini occupazionali: tali schemi, in particolare, sono l'oggetto dell'analisi che ho svolto in questo lavoro.

7.1.3 Contratti *Self – Enforcing*: riepilogo

Spesso è impossibile prevedere in un contratto tutte le contingente future e tutti gli aspetti che rilevano all'interno di una relazione economica. La frequente impossibilità di costruire un contratto esplicito ha creato notevoli problemi per ciò che riguarda il così detto *enforcement* contrattuale.

Fino ad ora è stato implicitamente ipotizzata la possibilità di ricorrere ad una corte per far valere in modo forzoso le clausole contrattuali (*enforcement* esogeno); tale possibilità, però, si basa strettamente sull'assunzione che il

contratto possa prevedere tutti gli aspetti salienti delle relazione economica. Questa eventualità è perlopiù remota, data la complessità delle dinamiche economico-sociali.

Nel momento in cui viene meno l'*enforcement* così detto esogeno, spetta, quindi, alle parti o al mercato creare strumenti finalizzati a far rispettare gli accordi. In altre parole, nell'impossibilità di ricorrere ad un giudice per far rispettare un contratto, sono necessari strumenti di *enforcement* endogeni, cioè nati all'interno della relazione economica o, comunque, all'interno del mercato. Il nuovo tipo di *enforcement* si divide in interni ed esterni.

Nel primo caso, l'*enforcing* è garantito tramite un sanzionamento rappresentato dalla cessazione dei flussi di quasi rendite positive nate all'interno della relazione economica stessa, sanzionamento posto in essere mediante l'interruzione del rapporto e il rifiuto di instaurare relazioni future con soggetti inadempienti. In questo caso la relazione deve avere durata ignota⁴ e deve produrre *surplus*; nel momento in cui queste due condizioni non possono essere rispettate (cosa molto frequente nella realtà), subentra l'*enforcement* endogeno, così detto, multilaterale.

Anche in questo caso il sanzionamento consiste nella cessazione dei flussi di quasi rendite positive, nate, però, non all'interno della relazione ma dalla permanenza dei soggetti sul mercato (o dall'affiliazione dei soggetti a particolari organizzazioni, come ad esempio quelle sindacali).

A questo punto, supponiamo che l'impresa, per garantirsi l'impegno dei lavoratori, prometta un bonus alla fine del periodo. Se il gioco è *one-shot*, ovviamente la promessa non è (generalmente) credibile: per lo più, infatti, il

⁴ Se la relazione avesse durata nota, l'*enforcement* bilaterale non avrebbe forza in quanto nell'ultimo periodo i soggetti si comporterebbero sempre in maniera opportunistica, pertanto la relazione economica non verrebbe instaurata a priori.

costo di breve periodo (pagamento del bonus) è maggiore del vantaggio che l'impresa può trarre (vantaggio che in un contesto uniperiodale è nullo).

Affinché la promessa di pagamento venga creduta, si rendono necessari sistemi reputazionali: solo all'interno di un contesto intertemporale è possibile considerare i *pay-off* attesi e quindi determinare sotto quali condizioni il costo di breve periodo (bonus) è maggiore del vantaggio di lungo periodo, rappresentato dall'incremento della produttività dovuto al consolidamento della reputazione.

7.1.4 *Rank-Order Tournaments*: riepilogo

L'incentivazione lavorativa tramite l'implementazione di un torneo, consiste nella valutazione delle sole *performance* relative dei lavoratori. A differenza di quello che sembrerebbe emergere dallo studio degli altri schemi di incentivazione, il rapporto lavorativo (quantomeno nell'ottica dell'impresa) non è mai puramente bilaterale; in ciascuna impresa esistono, infatti, soggetti che operano fianco a fianco e che hanno mansioni (*task*) simili.

Una classifica prestazionale relativa è, in ogni caso, di facile creazione, tuttavia, all'interno di una stessa impresa l'operato di ciascun lavoratore è, in larga misura, affetto dalle medesime variabili casuali. Questo aspetto, conferisce alla classifica relativa un'ulteriore punto di forza, consentendo di ottenere un livello superiore d'efficienza.

Ad ogni posizione in classifica è associato un certo livello salariale; nel modello base di riferimento (Lazear e Rosen, 1981), in particolare, viene ipotizzata la presenza di due soli soggetti, il “vincitore” riceve W , mentre il “secondo classificato” riceve w tale per cui $W > w$. Cioè che in realtà conta, però, non è tanto il livello salariale associato a ciascuna posizione in classifica, quanto il

differenziale ($W-w$) che, di fatto, rappresenta il vero e proprio premio e la variabile sulla quale i soggetti (imprese e lavoratori) prendono le decisioni.

I lavoratori, in particolare, profonderanno un maggior impegno lavorativo per vincere la competizione (la probabilità p di vittoria è, infatti, funzione, dell'*effort*). In assenza di particolari restrizioni, l'impresa può fissare liberamente lo *spread* in modo da giungere al *first-best*; questo avviene, in particolare, nell'ipotesi di neutralità al rischio da parte dei lavoratori.

Questa ipotesi, nonostante rappresenti una (forte) semplificazione delle realtà, consente di ottenere risultati significativi (e realistici) dal momento che i tornei escludono l'opportunismo d'impresa essendo i premi fissati a priori in modo chiaro ed immutabile.

Nel caso di avversione al rischio, il raggiungimento del *first-best* non è automatico: lo *spread* salariale deve essere contenuto all'interno di un certo *range* come forma di assicurazione implicita per i lavoratori; questo implica un vincolo che può non consentire l'ottenimento della situazione migliore in assoluto.

Da notare, infine, che in assenza di restrizioni, il raggiungimento del *first-best*, è del tutto indipendente dalle condizioni di mercato: W e w dipendono dal grado di concorrenzialità del mercato, ma ai fini dell'efficienza ciò che conta è il loro differenziale che, invece, ne è indipendente.

7.1.5 Effetti macroeconomici, conclusioni qualitative: riepilogo

Come risulta evidente, i vari schemi analizzati, pur essendo tutti volti alla risoluzione dell'opportunismo post contrattuale, nascono da contesti teorici differenti e a causa di questo, differiscono in modo abbastanza significativo per

quanto riguarda l'approccio a livello microeconomico. In particolare, queste profonde diversità si rispecchiamo anche in forti differenze negli effetti macroeconomici ed in particolare occupazionali.

Una volta delineati, a livello microeconomico, i vari schemi di incentivazione, l'ultima parte del lavoro ha riguardato l'analisi degli effetti macroeconomici. Già dall'analisi microeconomica sarebbe stato possibile fare alcune importanti osservazioni riguardanti gli effetti occupazionali, ma le profonde differenze strutturali di tali schemi non avrebbero consentito un confronto formale e sistematico.

Nell'ultima parte della tesi, sono stati riproposti i diversi schemi di incentivazione, formalizzati attraverso il medesimo strumento analitico (*asset value equation*), in modo da poter confrontare in modo chiaro, diretto ed esplicito i diversi risultati.

Di particolare rilevanza all'interno di questa analisi "comparativa" è l'ipotesi di perfetta osservabilità dell'*effort* da parte dell'impresa⁵ e quella di rendimenti decrescenti del lavoro: la prima ipotesi, in particolare, implica che, nel caso di un torneo, uno *shirker* non possa mai vincere la competizione. Inoltre i lavoratori sono ipotizzati omogenei e le imprese identiche, quindi, implementeranno il medesimo sistema di incentivazione e prenderanno le stesse scelte occupazionali.

Considerando il valore scontato atteso della vita lavorativa in caso di alto e basso impegno, in ogni schema, è stato possibile determinare, all'interno di ciascun filone di incentivazione, le condizioni di incentivo-compatibilità affinché vi sia elevato impegno da parte dei lavoratori. Tali condizioni implicano, all'interno dell'impresa, scelte differenti per ciò che riguarda il numero di soggetti da assumere.

⁵ La perfetta osservabilità non si estende ad una autorità esterna che, quindi, non può intervenire; si rendono pertanto necessari sistemi di enforcement endogeni.

In particolare, il modello di Shapiro – Stiglitz (ovviamente) fornisce i risultati “standard”: la disoccupazione involontaria è uno strumento di disciplina.

I contratti impliciti, invece, di per sé non implicano disoccupazione involontaria, tuttavia i problemi di credibilità (relativi al pagamento del bonus), rendono impossibile il raggiungimento del livello occupazionale “massimo”, in quanto a tale livello la produttività marginale del lavoro risulterebbe troppo bassa affinché il costo di breve periodo (pagamento del bonus) sia minore del vantaggio di lungo periodo (incremento delle produttività a seguito del consolidamento della reputazione).

I tornei, invece, date le condizioni di incentivo–compatibilità, consentono di ottenere i medesimi risultati che si avrebbero in presenza di *enforcement* esogeno, quanto, cioè, è possibile ottenere il *first–best*.

Questo lavoro ha, quindi, messo in luce un aspetto molto importante: la disoccupazione involontaria, contrariamente a quel che sembrava emergere dai salari d’efficienza, non è sempre necessaria per eliminare il *moral hazard*.

Ipotizzando, in particolare, che l’impresa possa conoscere in modo completo l’impegno dei lavoratori (nonostante risulti impossibile l’implementazione di un sistema di *enforcement* esogeno), il torneo consente di ottenere il *first–best*, a differenza dei contratti impliciti che generano disoccupazione anche se ad un livello inferiore rispetto ai salari d’efficienza.

In particolare, la disoccupazione creata nel caso dei contratti impliciti *self–enforcing* è dovuta solamente al problema della credibilità dell’impresa: se non ci fossero problemi di credibilità, anche i contratti impliciti consentirebbero (date le assunzioni di base) di ottenere il *first–best*.

Più in generale, possiamo affermare che in presenza di assunzioni non eccessivamente restrittive (dove, in particolare, l’impresa non può conoscere in modo perfetto l’impegno lavorativo e quindi uno *shirker* può vincere il torneo),

tutti gli schemi di incentivazione volti a eliminare il *moral hazard* a livello microeconomico, producono un livello più o meno alto di disoccupazione involontaria.

I tornei, tuttavia, consentono di ottenere un livello d'efficienza sociale maggiore rispetto a qualsiasi altro schema di incentivazione, implicando un livello di disoccupazione più basso sia dei contratti *self-enforcing*, sia dei salari d'efficienza.

BIBLIOGRAFIA

Akerlof A. G., *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, The Quarterly Journal of Economics, Vol. 84, no. 3, pp 488-500. 1970

Brucchi L., *Manuale di economia del lavoro*, Il mulino. 2001

Carmichael H.L., *Can Unemployment Be Involuntary? Comment*, American Economic Association, vol. 75(5), pp 1213-1214. 1985

Carmichael H.L., *Self-enforcing Contracts, Shirking, and Life Cycle Incentives*, Journal of Economic Perspectives, pp 65-83. 1989

Dornbusch R., Fischer S., *Macroeconomia*, il Mulino. 1995

Floro E.C., *Fondamenti teorici della rigidità' salariale nell'ambito dei "Non Market Clearing Models"*, CELPE (Centre of Labour Economics and Economic Policy), University of Salerno, Discussion Papers 79. 2004

Frackx L., D'Amato A., Brose I., *Multitask Rank Order Tournaments*, Economics Bulletin, vol 10, pp 1-10. 2004

Graffi A., *Il ruolo dell'informazione nella relazione tra concorrenza ed inefficienza-x: ambiguità teoriche della "product market competition"*, Liuc Papers n. 197, Serie Economia e Impresa, 51. 2006

Grillo M., *Informazione asimmetrica e incentivi: il problema del moral hazard*, Note didattiche, Facoltà di scienze politiche di Pisa. 1991

Holstrom B., *Moral hazard in team*, The Bell Journal of Economics, vol.3, no. 2, pp 324-340. 1982

Hawking W.S., *Breve storia del tempo – Dal Big bang ai buchi neri*, Rizzoli. 1989

Kreps D. M., *Corso di microeconomia*, Il Mulino. 1993

Lazear, E. P, S. Rosen, *Rank-Order Tournaments as Optimal Labor Contracts*, Journal of Political Economy, pp 841-864. 1981

Lin C., Yang C., *Tournaments as a worker discipline device*, Economic Inquiry 39, pp 644-650. 2001

Lin C., Yang C., *Tournaments, minimum wage, and money non-neutrality*, Journal of Macroeconomics, vol. 25(4), pp 481-489. 2003

Meccheri N., *Employment with Alternative Incentive Schemes when Effort is Not Verifiable*, CEIS, Labour, vol. 19(1), pp 55-80. 2003

Nicita A., Scoppa V., *Economia dei contratti*, Carocci editore. 2005

Pavanelli G., *Valore, distribuzione, moneta. Un profilo di storia del pensiero economico*, FrancoAngeli. 2003

Prendergast C., *The Provision of Incentives in Firms*, Journal of Economic Literature, pp 7-63. 1999

Romer D., *Advanced Macroeconomic*, MacGraw-Hill. 2006

Rosen S., *Implicit Contract Theory*, The International Library of Critical Writings in Economics, vol. 35. 1994

Ruiz M. L., *Mercato o contratto?*, Stato o mercato? Intervento pubblico e architettura dei mercati, pp 163-190, Pavia, vol. I. 2001

Salituro B., *La teoria dei contratti impliciti, la rigidità salariale e la disoccupazione involontaria: una sintesi dei risultati*, Economia Politica, pp 31-54. 1985

Scoppa V., *Contratti incompleti ed enforcement endogeno, Una rassegna della letteratura*, Economia Politica, pp 391-440. 2003

Shapiro C., Stiglitz J.E., *Equilibrium Unemployment as a Worker Discipline Device*, American Economic Review, pp 433-443. 1984

Smith A., *La ricchezza delle nazioni*, Newton Compton. 1776

Solow R. M., *Another Possible Source of Wage Stickiness*, Journal of Macroeconomics 1, pp 79-82. 1979

Yun J., *On the Efficiency of the Rank-Order Contract under Moral Hazard and Adverse Selection*, Journal of Labor economics, vol. 15, issue 3, pp 466-494. 1997

