



**KOÇ**  
**ÜNİVERSİTESİ**  
SUNA KIRAÇ LIBRARY



**KOÇ UNIVERSITY INSTITUTIONAL REPOSITORY**  
**PUBLICATION GUIDE**  
**WHICH VERSION TO SUBMIT**



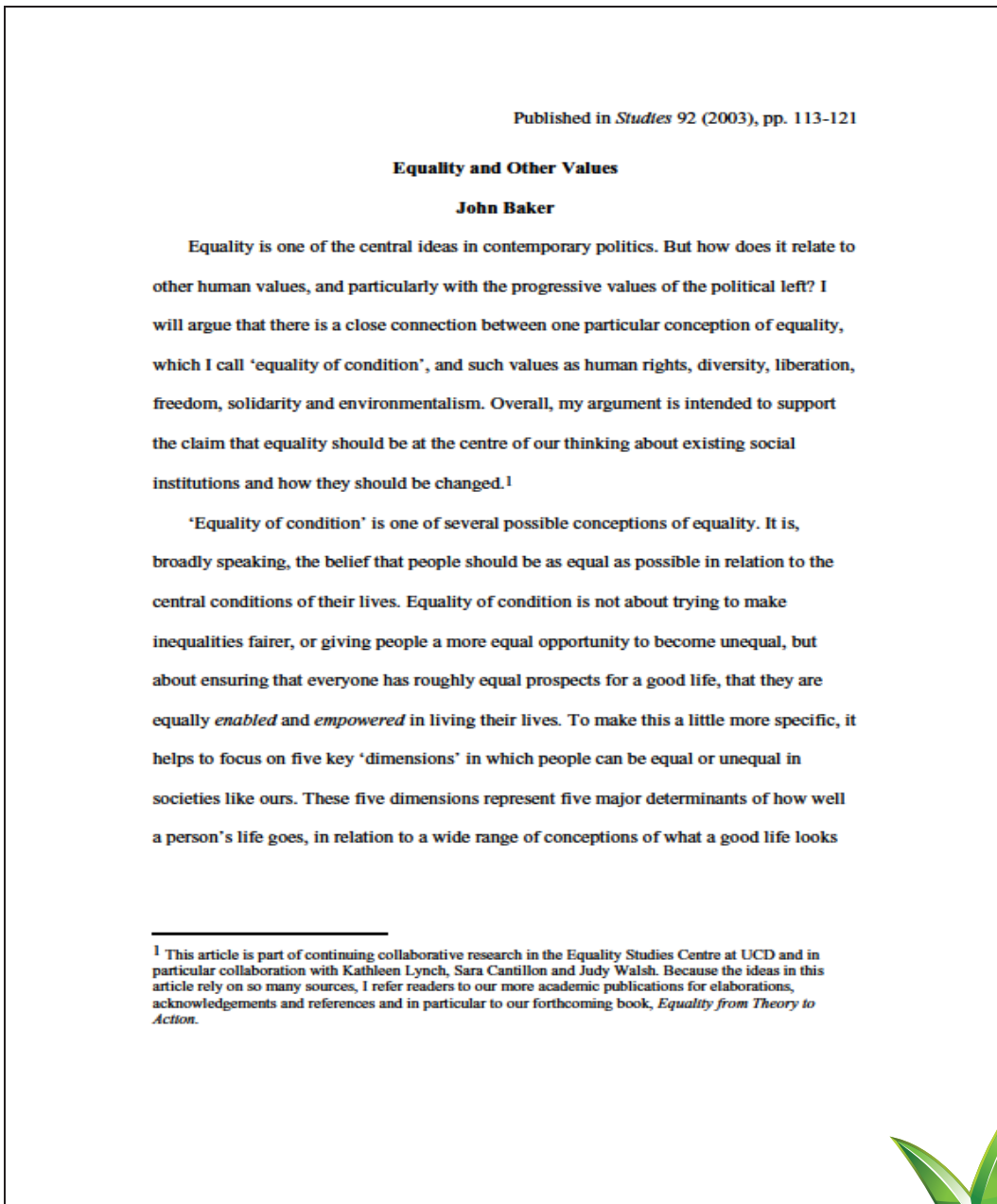
# Depositing full-text research to Koç University Institutional Repository

Visual examples of correct and incorrect versions

These examples illustrate the versions of items that we can and cannot accept into Koç University Institutional Repository.

**CORRECT: THE AUTHOR'S "POST-PRINT" OR AUTHOR'S FINAL VERSION OF THE ARTICLE AFTER THE REFEREEING PROCESS HAS TAKEN PLACE**

The following is what we are looking for: it is the unformatted author's version of the submitted or accepted paper:



## CORRECT: “CAMERA-READY”

The following is an example of a “camera-ready” paper where the publisher template is used, but the page numbering starts at 1 and there is no reference to the publication title:

---

### Tracking the Evolution of Communities in Dynamic Social Networks

---

**Derek Greene**  
School of Computer Science & Informatics  
University College Dublin  
derek.greene@ucd.ie

**Dónal Doyle**  
Idiro Technologies  
Dublin, Ireland  
donal.doyle@idiro.com

**Pádraig Cunningham**  
School of Computer Science & Informatics  
University College Dublin  
padraig.cunningham@ucd.ie

#### Abstract

Real-world social networks from a variety of domains can naturally be modelled as dynamic graphs. However, approaches to detecting communities have largely focused on identifying communities in static graphs. Recently, researchers have begun to consider the problem of tracking the evolution of groups of users in dynamic scenarios. Here we describe a model for tracking the progress of communities over time in a dynamic network, where each community is characterised by a series of significant evolutionary events. This model is used to motivate a community-matching strategy for efficiently identifying and tracking dynamic communities. Evaluations on synthetic graphs containing embedded events demonstrate that this strategy can successfully track communities over time in volatile networks. In addition, we describe experiments exploring the dynamic communities detected in a real mobile operator network containing millions of users.

#### 1 Introduction

Social network analysis methods have traditionally focused on the representation of graphs as static networks. This has been the case for the task of community detection, where the goal is to identify meaningful group structures in the network. However, by representing a dynamic source of data as a static network, group structures present over shorter periods of time can be difficult to identify or may be completely ablated. In addition, by discarding temporal information, the detail of the evolutionary behaviour of these groups is lost.

Modelling structural changes in networks is important in a wide range of real-world social network analysis problems, where the data naturally has a temporal aspect. The evolving nature of social media makes it a candidate for this type of analysis. Researchers may be interested in examining the formation and change in communities – such as clusters of frequently interacting authors in the blogosphere [13], or the formation of circles of friends in online networks such as Facebook and Twitter. Other application areas include the analysis of the evolution of research communities within and across academic disciplines [14]. A particularly relevant application is the analysis of mobile subscriber networks [17], where the activity of groups of users over time are potential predictors of future behaviour that is of specific interest to network operators, such as subscriber churn or handset adoption. However, the scale of such networks presents a challenge even for existing static community finding techniques.


1



## CORRECT: NUMBERED LINES

This version, which includes numbered lines and no publisher branding, is acceptable:

1  
2  
3  
4           **Background**  
5  
6           The peak unit shaft resistance ( $\tau_f$ ) mobilised by a pile in sand can be estimated using  
7  
8           earth pressure theory as:  
9  
10  
11  
12  
13           [1]                    $\tau_f = K \sigma'_v \tan \delta_f$   
14  
15  
16  
17  
18           where  $K$  is the earth pressure coefficient,  $\sigma'_v$  is the in-situ vertical effective stress and  
19  
20            $\delta_f$  is the soil-pile interface friction angle. A common difficulty with the application of  
21  
22           Equation 1 is the choice of an appropriate  $K$  value for design. Paikowsky (2004) notes  
23  
24           that design methods proposed by Reese and O'Neill (1999) are in widespread use.  
25  
26           They suggest  $K/K_0$  (where  $K_0$  is the coefficient of earth pressure at rest), varies with  
27  
28           the pile construction method, varying from 0.67 when the pile is excavated using  
29  
30           slurry, to 1.0 in a dry excavation.  $K_0$  is notoriously difficult to measure but can be  
31  
32           estimated using the method proposed by Mayne and Kulhawy (1982):  
33  
34  
35  
36  
37  
38  
39                            $K_0 = (1 - \sin \phi'_p)$     for normally consolidated soil  
40           [2]                            $K_0 = (1 - \sin \phi'_p) OCR^{\sin \phi'_p}$    for overconsolidated soil  
41  
42  
43  
44  
45  
46           where  $\phi'_p$  is the peak friction angle and  $OCR$  is the Over-Consolidation Ratio.  
47  
48  
49           Where accurate estimates of  $K_0$  are unavailable, Reese and O'Neill suggest an  
50  
51           empirical correlation based on a conservative estimate of the shaft resistance (in kPa)  
52  
53           measured from a series of field tests:  
54  
55  
56  
57  
58           [3a]                    $\tau_f = \beta \sigma'_v$   
59  
60  
61  
62  
63  
64  
65



## INCORRECT: PUBLISHER'S PDF

In the majority of cases, the publisher's PDF cannot be accepted. Published versions will contain pagination and other publisher-specific types of data (copyright symbol, DOI, branding), for example:



# Feature Selection for Gene Expression Using Model-Based Entropy

Shenghuo Zhu, Dingding Wang, Kai Yu, Tao Li, and Yihong Gong

**Abstract**—Gene expression data usually contain a large number of genes but a small number of samples. Feature selection for gene expression data aims at finding a set of genes that best discriminate biological samples of different types. Using machine learning techniques, traditional gene selection based on empirical mutual information suffers the data sparseness issue due to the small number of samples. To overcome the sparseness issue, we propose a model-based approach to estimate the entropy of class variables on the model, instead of on the data themselves. Here, we use multivariate normal distributions to fit the data, because multivariate normal distributions have maximum entropy among all real-valued distributions with a specified mean and standard deviation and are widely used to approximate various distributions. Given that the data follow a multivariate normal distribution, since the conditional distribution of class variables given the selected features is a normal distribution, its entropy can be computed with the log-determinant of its covariance matrix. Because of the large number of genes, the computation of all possible log-determinants is not efficient. We propose several algorithms to largely reduce the computational cost. The experiments on seven gene data sets and the comparison with other five approaches show the accuracy of the multivariate Gaussian generative model for feature selection, and the efficiency of our algorithms.

**Index Terms**—Feature selection, multivariate Gaussian generative model, entropy.

## 1 INTRODUCTION

GENE expression refers to the level of production of protein molecules defined by a gene. Monitoring of gene expression is one of the most fundamental approach in genetics and molecular biology. The standard technique for measuring gene expression is to measure the mRNA instead of proteins, because mRNA sequences hybridize with their complementary RNA or DNA sequences while this property lacks in proteins. The DNA arrays, pioneered in [5] and [10], are novel technologies that are designed to measure gene expression of tens of thousands of genes in a single experiment. The ability of measuring gene expression for a very large number of genes, covering the entire genome for some small organisms, raises the issue of characterizing cells in terms of gene expression, that is, using gene expression to determine the fate and functions of the cells. The most fundamental of the characterization problem is that of identifying a set of genes and its expression patterns that either characterize a certain cell state or predict a certain cell state in the future [18].

When the expression data set contains multiple classes, the problem of classifying samples according to their gene expression becomes much more challenging, especially when the number of classes exceeds five [24]. Moreover, the special characteristics of expression data add more

challenge to the classification problem. Expression data usually contain a large number of genes (in thousands) and a small number of experiments (in dozens). In machine learning terminology, these data sets are usually of very high dimensions with undersized samples. In microarray data analysis, many gene selection methods have been proposed to reduce the data dimensionality [31].

Gene selection aims to find a set of genes that best discriminate biological samples of different types. The selected genes are “biomarkers,” and they form a “marker panel” for analysis. Most gene selection schemes are based on binary discrimination using rank-based schemes [8] such as information gain, which reduces the entropy of the class variables given the selected features. One critical issue in these rank-based methods is data sparseness. For example, the estimation of the traditional information gain is an empirical estimation directly on the data. Suppose we select the 11th gene for a data set. The 10 selected genes split the training data into  $1,024 = 2^{10}$  groups (assuming that each gene does a binary split). Since we have very few samples in most groups, the estimations of mutual information between the 11th gene and the target in each group are not accurate. Thus, the information gain, which is the sum of the mutual information over all groups, is not accurate.

To overcome the issue of data sparseness, we propose a model-based approach to estimate the entropy on the model, instead of on the data themselves. Here, we use multivariate Gaussian generative models, which model the data with multivariate normal distributions. Multivariate normal distributions are widely used in various areas, including gene expression data [34], because of their *generality* and *simplicity*. The means of variables (expression data of genes and class labels) and the covariances between them are two basic measures of variables themselves

- S. Zhu, K. Yu, and Y. Gong are with NEC Laboratories America, 10080 North Wolfe Road, Suite SW3-350, Cupertino, CA 95014. E-mail: {zsh, ky, ygong}@research.nec-labs.com.
- D. Wang and T. Li are with the School of Computer Science, Florida International University, 11200 SW 8th Street, Miami, FL 33199. E-mail: {dawang003, taoli}@cs.fiu.edu.


Manuscript received 18 Oct. 2007; revised 17 Jan. 2008; accepted 13 Mar. 2008; published online 10 Apr. 2008.  
For information on obtaining reprints of this article, please send e-mail to: [tcb@computer.org](mailto:tcb@computer.org), and reference IEEECS Log Number TCB-2007-10-0136. Digital Object Identifier no. 10.1109/TCBB.2008.35.



## INCORRECT: PUBLISHER'S PDF DOWNLOADED FROM PROVIDER

The following example is both in breach of copyright (as the published version) and also the provider's licence agreement:

This article was downloaded by: [IRIS]  
On: 12 April 2010  
Access details: Access Details: [subscription number 920933949]  
Publisher Routledge  
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



PROMETHEUS  
critical studies in innovation

Routledge  
Taylor & Francis Group

**Prometheus**  
Publication details, including instructions for authors and subscription information:  
<http://www.informaworld.com/smpp/title-content=t713442842>

**The immediate practical implication of the Houghton Report: provide Green open access now**  
Stevan Harnad<sup>a,b</sup>  
<sup>a</sup> Institut des Sciences Cognitives (ISC), Université du Québec à Montréal, Montréal, Québec, Canada  
<sup>b</sup> Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton, UK

Online publication date: 24 March 2010

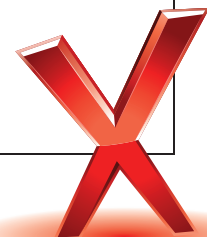
To cite this Article Harnad, Stevan (2010) 'The immediate practical implication of the Houghton Report: provide Green open access now', *Prometheus*, 28: 1, 55 – 59  
To link to this Article: DOI: 10.1080/08109021003676367  
URL: <http://dx.doi.org/10.1080/08109021003676367>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

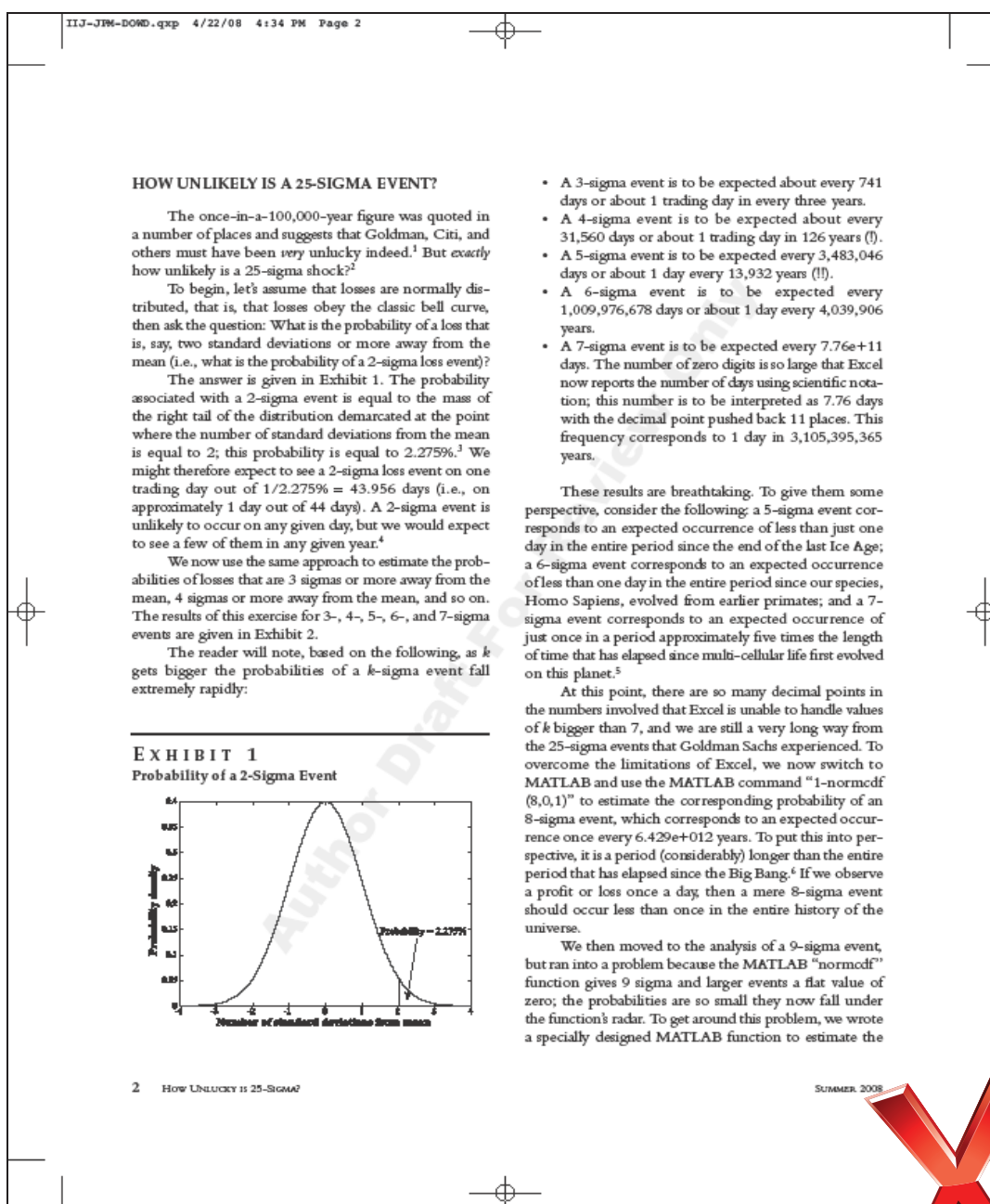
This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.



# INCORRECT: WATERMARKED DOCUMENTS CONTAINING PUBLISHER'S TYPESETTING & OTHER SIGNS OF PUBLISHER PROCESSING

Below are some examples of watermarked docs we **cannot** use as they contain the publisher typesetting and other signs of publisher processing, etc: "Uncorrected proof", "Author proof", "Author Draft – for review only", "Author personal copy":





Available online at [www.elsevier.com/locate/jirfa](http://www.elsevier.com/locate/jirfa)

SCIENCE @ DIRECT®

International Review of Financial Analysis  
xx (2004) xxx–xxxIRFA  
INTERNATIONAL REVIEW OF  
FINANCIAL ANALYSIS

## International equity market integration in a small open economy: Ireland January 1990–December 2000

John Cotter\* 4

*Department of Banking and Finance, Centre for Financial Markets, University College Dublin,  
Graduate School of Business, Blackrock, Co. Dublin, Ireland*

---

**Abstract** 8

We examine the relationship between the Irish, German, UK and U.S. equity markets. Our main finding is that the Irish equity market depends heavily on trading activity in the other markets but not vice versa. Significant return and volatility spillover effects occur in the direction of, but not from, the Irish market. We also find that dual listing in the form of American Depositary Receipts (ADRs) has an important role to play in these spillover effects. Our findings obtain throughout the sample, but are strongest for the period after the ERM crises and before the introduction of the euro.

© 2004 Published by Elsevier Inc. 15

*JEL classification:* G1; G10 17*Keywords:* Equity market linkages; Return and volatility spillovers; Dual trading 18

---

**1. Introduction** 19

Ireland is recognized as a small open economy with a heavy reliance on external trade that has been increasing over time (EUROSTAT, 2000). The nature of Ireland's capital flows is less clear, however, and this paper addresses this subject by describing the bivariate interactions between the Irish equity market and the markets in Germany, the United Kingdom and the United States. Traditional strong interactions between the Irish and the UK markets for economic and political purposes may have been superseded by

\* Tel: +353-1-716-8900; fax: +353-1-283-5482.

E-mail address: [john.cotter@ucd.ie](mailto:john.cotter@ucd.ie) (J. Cotter).

## REEVALUATING HEDGING PERFORMANCE

JOHN COTTER\*  
JIM HANLY

Mixed results have been documented for the performance of hedging strategies with the use of futures. This article reinvestigates this issue with the use of an extensive set of performance-evaluation metrics across seven international markets. The hedging performances of short and long hedgers are compared with the use of traditional variance-based approaches together with modern risk-management techniques, including value at risk, conditional value at risk, and approaches based on downside risk. The findings indicate that use of these metrics to evaluate hedging performance yields differences in terms of best hedging strategy as compared with the traditional variance measure. Also, significant differences in performance between short and long hedgers are found. These results are observed both in sample and out of sample. © 2006 Wiley Periodicals, Inc. Jrl Fut Mark 26:000–000, 2006

The authors would like to thank an anonymous referee and the editor, Bob Webb, for their constructive comments. Cotter's contribution to the study has been supported by a University College Dublin Faculty of Commerce research grant.

\*Correspondence author, Director of Centre for Financial Markets, Department of Banking and Finance, University College Dublin, Blackrock, Co. Dublin, Ireland; e-mail: john.cotter@ucd.ie.

Received April 2005; accepted July 2005

- John Cotter is the Director of the Centre for Financial Markets in the Department of Banking and Finance at University College Dublin in Dublin, Ireland.
- Jim Hanly is at the School of Accounting and Finance at Dublin Institute of Technology in Dublin, Ireland.

The Journal of Futures Markets, Vol. 26, No. 5, 000–000 (2006) © 2006  
Wiley Periodicals, Inc.  
Published online in Wiley InterScience (www.interscience.com)  
DOI: 10.1002/fut.20212

WILEY  
InterScience®  
DISCOVER SOMETHING GREAT





Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

 ScienceDirect

Economics Letters 94 (2007) 83–89

**economics  
letters**

[www.elsevier.com/locate/econbase](http://www.elsevier.com/locate/econbase)

## Dealing with monopsony power: Employment subsidies vs. minimum wages

Eric Strobl, Frank Walsh \*

*Ecole Polytechnique and University College Dublin, Ireland*

Received 24 November 2003; received in revised form 16 June 2006; accepted 11 August 2006  
Available online 14 November 2006

### Abstract

We show in a monopsony model that accounting for changes in hours a minimum wage has ambiguous effects on employment and welfare. When all workers have the same preference ordering over leisure and consumption employment subsidies unambiguously improve welfare. Many countries have minimum wages and also tax minimum wage workers.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Monopsony; Minimum wages; Hours worked

*JEL classification:* J42; J48

### 1. Introduction

Manning (2003) has argued for the importance of monopsony power in modern labour markets.<sup>1</sup> Importantly, monopsony models result in inefficient market outcomes that provide a rationale for regulation. Minimum wages tend to be the common policy response while minimum wage workers are often taxed. In this paper we demonstrate that minimum wages may lower employment and welfare in a monopsony model while subsidies rather than taxes are efficient for a common class of utility functions.

\* Corresponding author, Economics Department, g216, John Henry Newman building, University College Dublin, Dublin 4 Ireland. Tel.: +353 1 7168697; fax: +353 1 2830068.

*E-mail address:* [Frank.Walsh@ucd.ie](mailto:Frank.Walsh@ucd.ie) (F. Walsh).

<sup>1</sup> Manning (2003) p.360 lists the sources of monopsony power as "Ignorance among workers about labor market opportunities, individual heterogeneity in preferences over jobs and mobility costs".

0165-1765/\$ - see front matter © 2006 Elsevier B.V. All rights reserved.  
doi:10.1016/j.econlet.2006.08.005



## INCORRECT: AUTHOR'S PERSONAL COPY

The previous example illustrates an “Author’s personal copy” through the use of a watermark. This example illustrates another form of an author’s personal copy:

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

