

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.

| DATE DUE | DATE DUE | DATE DUE |
|-------------|----------|----------|
| FEB 05 1997 | | |
| JAN 12 2009 | | |
| 07 02 08 | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

MSU is An Affirmative Action/Equal Opportunity Institution

c:\circ\datedue.pm3-p.1

THREE ESSAYS ON SHARE CONTRACTS,
LABOR SUPPLY, AND THE ESTIMATION OF
MODELS FOR DYNAMIC PANEL DATA

By

Seung Chan Ahn

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1990

ABSTRACT

THREE ESSAYS ON SHARE CONTRACTS, LABOR SUPPLY, AND THE ESTIMATION OF MODELS FOR DYNAMIC PANEL DATA

By

Seung Chan Ahn

This dissertation deals with three topics: share contracts, labor supply, and the estimation of models for dynamic panel data. Chapter 1 proposes a model which predicts under generally acceptable assumptions, fixed wages across different economic states and lay-offs for bad states, and shows that a share contract exists that Pareto-dominates and has no less employment than the fixed-wage contract. Chapter 2 considers joint estimation of the determinants of the employment status of married women, their labor-force participation decisions, and their market wages. The empirical results imply that recognizing frictions in the labor market is important to explain the determinants of individuals' employment status in a concrete and correct way. The estimation procedure including the wage equation generates more significant and reasonably signed estimates. Chapter 3 considers a dynamic model using panel data which include a large number of cross-section observations, but only over a short period of time. This chapter proposes an estimator that is efficient under general circumstances.

TABLE OF CONTENTS

| | |
|-----------------------------------------------------------------------------------------------------------|-----------|
| Introduction | 1 |
| Chapter 1 | |
| A share Economy as a Work incentive Device | 4 |
| I. Introduction | 4 |
| II. Wage Contract Model | 9 |
| III. Share Contract Model | 17 |
| IV. Conclusion | 24 |
| Appendix | 29 |
| References | 32 |
| Chapter 2 | |
| The Joint Estimation of a Model of Labor Force and Employment Decisions and Market Wages . . . | 33 |
| I. Introduction | 33 |
| II. Model | 36 |
| III. Data | 50 |
| IV. Empirical Results | 53 |
| V. Conclusion | 64 |
| Appendix A | 75 |
| Appendix B | 78 |
| Appendix C | 81 |
| Appendix D | 86 |
| References | 89 |

Table of Contents (cont'd)

Chapter 3

Efficient Estimation of Models for Dynamic Panel Data

| | |
|--------------------------------------------------|-----|
| I. Introduction | 91 |
| II. Conventional IV Methods | 94 |
| III. Derivation of Moment Conditions | 98 |
| IV. Estimation | 108 |
| V. Estimation with Exogenous Variables | 115 |
| VI. Conclusion | 124 |

| | |
|----------------------|-----|
| Appendix A | 127 |
| Appendix B | 130 |
| Appendix C | 132 |
| Appendix D | 146 |
| Appendix E | 148 |
| Appendix F | 151 |

| | |
|----------------------|-----|
| References | 154 |
|----------------------|-----|

LIST OF TABLES

| | |
|---------------------|-----|
| Table 1 | 66 |
| Table 2 | 67 |
| Table 3 | 68 |
| Table 4-A | 69 |
| Table 4-B | 69 |
| Table 5 | 70 |
| Table 6 | 71 |
| Table 7 | 72 |
| Table 8 | 73 |
| Table 9 | 74 |
| Table 10 | 103 |

INTRODUCTION

This dissertation deals with three topics: share contracts, labor supply, and the estimation of models for dynamic panel data. Since each topic is independent of the others, each of following chapters focuses on one topic and contains its own introduction and conclusion sections.

Chapter 1 attempts to provide a theoretical basis for share contracts. There are many studies comparing share and fixed-wage contracts from the point of view of welfare or/and employment. However, their methods of comparison are arbitrary, in the sense that they simply assume the fixed-wage contract to be optimal among wage contracts. A better comparison could be done by investigating the conditions which generate fixed wages across different economic states, and examining whether a share contract could perform better than a fixed-wage contract under those conditions. For this reason, I propose a microeconomic model which predicts, under generally acceptable assumptions, fixed wages across different economic states and lay-offs for bad states. I then show that a share contract exists that Pareto-dominates and has no less employment than the fixed-wage contract. This result implies that share contracts could not only improve every economic

agent's well-being, but also stabilize the employment level in the economy.

Chapter 2 considers joint estimation of the determinants of the employment status of married women, their labor-force participation decisions, and their market wages.

Many of the previous studies of labor supply assume that the employment status of an individual is determined solely by his/her desire to work. Those studies treat the unemployed and non-participants as behaviorally equivalent, ignoring frictions in the labor market. In this chapter, the unemployed are regarded as willing to work but not successful in their job search, and therefore they are treated as behaviorally different from non-participants. Therefore, the model considered in this chapter consists of two equations describing employment and labor-force decisions, and also a wage equation. The empirical results given in this chapter imply that recognizing frictions in the labor market is important to explain the determinants of individuals' employment status in a concrete and correct way, and that the traditional labor-supply model generates biased estimates of the determinants of willingness to work. Furthermore, compared to other methods for joint estimation of labor-force and employment decisions, the estimation procedure including the wage equation generates more significant and reasonably signed estimates. Significant sample selection biases generated by employment and participation decisions are also detected in the distribution of observed wage rates, and they

are successfully corrected by the joint estimation procedure.

Chapter 3 considers a dynamic model using panel data which include a large number of cross-section observations, but only over a short period of time. This chapter proposes an estimator that is efficient under general circumstance. Several authors have proposed simple but consistent instrumental-variable (IV) estimators, which are identical to generalized-methods-of-moments (GMM) estimators based on some available moment conditions. GMM estimators are efficient in general circumstances if all known is a certain set of moment restrictions. This chapter adopts standard assumptions for the dynamic panel data model, and characterizes all of the moment conditions that these assumptions imply. It turns out that previous studies do not impose all of the available moment conditions, which reveals the inefficiency of the previous IV estimators. The estimator proposed in this chapter is efficient because it is obtained exploiting all useful information from the standard assumptions.

Chapter 1

A Share Economy as a Work Incentive Device

I. Introduction

Stagflation during the last two decades has put an end to the *Golden era* of the Keynesian doctrine. If the Phillips curve is downward sloping, then a trade-off between unemployment and inflation must exist, and the government can choose a desirable combination of them. However, the lesson we have learned during last 20 years is that the long-run Phillips curve seems to be vertical. High unemployment and inflation do not alternate; they rather frequently occur simultaneously. Government is no longer able to buy employment at the cost of inflation. In a Keynesian framework, an expansionary policy successfully reduces the unemployment rate, since inflation drives down real wages. However, the public's expectation of inflation seems to catch up to actual inflation so quickly that nominal wages increase at approximately the same rate as the general price level. Therefore, government's expansionary policies often create high inflation without affecting employment even during recessions. This means that real as well as nominal wages seem to be rigid.

Through a series of publications, M. Weitzman argues that stagflation is tied to wage rigidity. He argues that "stagflation is just an unfortunate consequence of the wage-payment system."¹ His basic view is identical to the

Keynesians' in that he believes that all problems have their origins in wage rigidity. However, his prescription differs from the Keynesians' in that he emphasizes the necessity of making wages flexible. His suggestion is to tie compensation to "an appropriate index of the firm's performance, say a share of its revenues or profits."² For simplicity, imagine a payment system in which the wage rate is determined by two components -- some fixed compensation and a portion depending on firm's total revenue. In this case, wages will move in the same direction as the firm's performance. By a simple algebraic equation, Weitzman shows that firms are always characterized by an excess demand for labor. Therefore, in a share economy, firms behave like a *vacuum cleaner*, constantly searching for employees and eagerly sucking up all the unemployed. In a share economy, capitalism not only guarantees Consumer Sovereignty; but Worker Sovereignty.³ With this belief, Weitzman suggests that the government should offer tax incentives in order to get firms to adopt share contracts.

There are two main criticisms of Weitzman's analysis. His ideas can be summarized in two propositions. First, in a share economy compensation is no longer rigid, and it adjusts in a manner that leads the economy to full employment even in a short run. Second, the share economy improves the welfare of economic agents. Nordhaus [1988] and John [1987] have criticized Weitzman's first proposition. Nordhaus argues that Weitzman's analysis of the short-run behavior of a share

economy omits a detailed specification of labor supply, and shows that the excess-demand proposition no longer holds when labor supply constraints are introduced. John shows that a share economy may actually lead to greater employment fluctuation depending on the specification of labor supply curves. Share contracts will have less employment fluctuation only if the share parameters are determined by correct information about the demand for and supply of labor.

Cooper, (see Nordhaus and John [1986]) amongst others, has criticized the second proposition. Implicit contract and efficiency wage theories provide some intuition on the rigid wage phenomena. The former suggests that risk-averse workers would prefer rigid wages. The latter argues that wages are rigid downward in order to prevent workers from shirking. If wage rigidity comes from the self-interests of economic agents, then the payment system will not allow wages to fluctuate. Weitzman does not deny that fixed wages would be optimal at a microeconomic level. His basic assertion is that the wage contracts are not optimal at a macroeconomic level. In order to support this claim, Weitzman [1985] shows that share contracts could increase employment while at the same time offering approximately the same compensation as under fixed-wage contracts. This result is based on a macroeconomic model. In response, Cooper [1988] shows that an injection of share contracts into one sector of a two-sector economy will yield a Pareto-improving resource allocation only in a special case. That is, a share economy, in which all firms adopt

share contracts, may be superior to a wage economy, but share contracts adopted by a subset of sectors may not help the whole economy.

This paper is an attempt to provide support for the share economy, which Wietzman's model fails to do. If there exists a share contract that Pareto-dominates, and has no less employment than a wage contract in a microeconomic model, then the two main criticisms described above will be no longer valid. The Pareto-dominating share contract will be able to help the whole economy without suspending some group's self-interest. Employment will also fluctuate less in a share economy than in a wage economy. I demonstrate this in two steps.

Some implicit contract models predict constant compensation across states of nature under fairly strong assumptions. My first focus is on whether constant compensation could be observed under more general conditions, in particular, when it is necessary to monitor labor's effort on the job. I assume only risk-neutral firms and risk-averse workers. Workers may have an incentive to shirk on the job, once they are employed. Shapiro and Stiglitz's [1984] efficiency wage model shows how the monitoring cost on workers could bring about downward wage rigidity and unemployment. My key point is the introduction of a no-shirking condition (NSC) to the implicit contract framework as an incentive compatible mechanism. My model predicts fixed compensation across states with lay-offs in bad states. If there is no monitoring cost,

firms maintain a higher employment level even in bad states to insure risk-averse workers. If there is a monitoring cost, firms have an incentive to decrease the employment level to save on monitoring. They cannot easily cut compensation for workers, because lower wages give the employed an incentive to shirk. This intuition partly explains my results.

As a second step, I compare share contracts with the optimal wage contract. Not surprisingly, there always exists a share contract which Pareto-dominates the optimal wage contract. The reason is quite simple. Under a wage contract, shirkers still can get the agreed compensation regardless of whether or not they shirk. On the contrary, under a share contract, shirkers, themselves, suffer from their shirking, because the firm's total output and revenue decrease when shirking occurs. Therefore, firms can reduce the monitoring cost per worker due to workers' decreased incentive to shirk. Furthermore, I show the existence of some forms of share contracts which have no less employment than the wage contract.

My model has a very different implication from Weitzman's. The share contracts in Weitzman's model Pareto-dominate the wage contracts at a macroeconomic level, not at a microeconomic level. The share contracts have a positive effect on the economy-wide employment level, which in turn increases aggregate demand, improving all firms' market conditions. The workers who are already employed will initially suffer from lower wages. However, their firms'

improved profitability under a share economy finally will compensate their suffering with higher wages. This reasoning is correct only when sufficiently many firms adopt share contracts. Instead my model predicts Pareto-dominance of share contracts over wage contract by a different mechanism. Share contract creates some cost to shirkers. Monitoring cost, a pure social welfare loss, will decrease. We can also choose a share contract that has no less employment than the wage contract has. Higher welfare and employment will result. This is possible even at a micro level.

Section II shows that the optimal wage contract has fixed wages across the states. Section III proves that a share contract exists that Pareto-dominates the optimal wage contract, and has no less employment. Section IV summarizes some conclusions.

II. Wage Contract Model

First, consider a labor contract between a firm and N homogeneous workers. For simplicity, assume that the firm uses only labor to produce a single commodity. Output depends on total employment (E) and each worker's level of effort (e), which are perfect substitutes in the production process. Define the revenue function by $sf(eE)$, where s denotes an predictable product-demand shock. Here, $f(\cdot)$ is strictly increasing and strictly concave, i.e., $f' > 0$, $f'' < 0$. For simplicity, I assume that each employee's working hours are fixed for technological reason. We can relax this assumption

without changing the results in this section. (For details, see APPENDIX) Then, we have the profit function:

$$(1) \quad \pi(s) = sf(e(s)E(s)) - w^e(s)E(s) - w^u(s)(N-E(s))$$

where w^e is the wage paid to each employed worker and w^u is the severance pay for the unemployed.

Each worker has the same concave utility function:

$$(2) \quad U = U(Y, e)$$

where Y denotes consumption. Assume that $U_y > 0$, $U_e < 0$. For employed workers, consumption in state s is given by $w^e(s)$, and for the unemployed, $w^u(s)$. Assume that the firm is risk-neutral, and workers are risk-averse. Then, δ_w describes a wage contract with

$$\delta_w = \{E(s), e(s), w^e(s), w^u(s)\}$$

The optimal contract, δ_w^* , can be characterized by the solution to the problem:

$$(C.1) \quad \max_{\delta} E_s \pi(s)$$

subject to

$$E_s \{ (E(s)/N) U(w^e(s), e(s)) + (1-E(s)/N) U(w^u(s), 0) \} \geq U_0$$

$$0 \leq E/N \leq 1$$

$$e \geq 0, \quad \text{for all } s$$

where U_0 is a utility level of a worker obtained in the worker's next best alternative. The first constraint will be binding, because otherwise the firm could lower w^e and increase profit. (See Cooper [1987].) This formulation is very close to Cooper's [1987] basic implicit contract model. The only difference is that I use e rather than the worker's hours as in Cooper's model. This difference makes my model

similar to the *principal-agent* model [1978]. Basically, *principal-agent* models are designed to show how a firm could improve its workers' productivity with some specific compensation scheme. As efficiency wage models suggest, worker productivity will be related to the wage rate. The model specified above offers the mechanism generating correlation between productivity and the wage in an implicit contract framework.

The first interesting result that arises from (C.1) is that for any s , the firm employs all workers --- there is full employment. We can summarize this as following.

PROPOSITION 1. In any optimal contract, $E(s)=N$ for all s .

Proof. Suppose not, i.e., in δ_w^* , there exists a state, s_1 , with $E(s_1) < N$. Given s_1 , a worker's expected utility is given by

$$(E(s_1)/N)U(w^e(s_1), e(s_1)) + (1-E(s_1)/N)U(w^e(s_1), 0)$$

Consider another contract, $\underline{\delta}_w$, such that $\underline{\delta}_w$ is identical to δ_w^* , for all states other than s_1 , and at s_1 , there is full employment with

$$\underline{e}(s_1)N = e(s_1)E(s_1)$$

$$\underline{w}^e(s_1) = (E(s_1)/N)w^e(s) + (1-E(s_1)/N)w^u(s_1)$$

$$\underline{w}^u(s_1) = 0$$

Then,

$$\begin{aligned}\pi(s_1) &= s_1 f(\underline{e}(s_1)N) - \underline{w}^e(s_1)N \\ &= s_1 f(e(s_1)E(s_1)) - (E(s_1))w^e(s_1) - (N-E(s_1))w^u(s_1)\end{aligned}$$

$$= \pi^*(s_1)$$

Therefore, the firm is indifferent between $\underline{\delta}_w$ and δ_w^* . Now, compare the worker's expected utility under the two contracts:

$$\begin{aligned} & (E(s_1)/N)U(w^e(s_1), e(s_1)) + (1-E(s_1)/N)U(w^e(s_1), 0) \\ & < U((E(s_1)/N)w^e(s_1) + (1-E(s_1)/N)w^e(s_1), E(s_1)e(s_1)/N) \\ & = U(\underline{w}^e(s_1), \underline{e}(s_1)) \end{aligned}$$

The inequality is due to the assumption that workers are risk-averse. Thus, workers prefers $\underline{\delta}_w$ to δ_w^* , so that δ_w^* is not an optimal contract.

QED

(C.1) implicitly assumes that the contract can be enforced voluntarily. However, this assumption is quite unrealistic. Even when a worker shirks, he still gets the agreed compensation. Since all workers are identical, no one will work. Therefore, the firm has to monitor workers in order to sort out those who shirk on the job. Assume that the firm bears some cost (C) when monitoring a worker. Let m be the probability of catching a given shirker. Assume

$$(3) \quad C = C(m); \quad C' > 0$$

Now, each employed worker decides whether or not to shirk. If an employed worker does not shirk, he gets utility of $U(w^e, e)$. Otherwise he gets expected utility of $(1-m)U(w^e, 0) + mU(0, 0)$, because shirkers, once caught, are fired immediately. To prevent workers from shirking, the following condition (no-shirking condition; NSC) must be satisfied:

$$(4) \quad U(w^e, e) \geq (1-m)U(w^e, 0) + mU(0, 0)$$

for all s . In this case, each firm faces the following profit function:

$$(5) \quad \pi(s) = sf(e(s)E(s)) - w^e(s)E(s) - w^u(s)(N-E(s)) \\ - C(m(s))E(s)$$

Denote a contract by

$$\delta_w = \{E(s), e(s), w^e(s), w^u(s), m(s)\}$$

Then, the optimal contract, δ_w^* , solves

$$(C.2) \quad \max_{\delta} E_s \pi(s)$$

subject to

$$(C.2-1) \quad E_s \{ (E(s)/N)U(w^e(s), e(s)) + (1-E(s)/N)U(w^u(s), 0) \} \geq U_0$$

$$(C.2-2) \quad U(w^e(s), e(s)) \geq (1-m(s))U(w^e(s), 0) + m(s)U(0, 0)$$

$$(C.2-3) \quad 0 \leq E(s)/N \leq 1$$

The second constraint will bind, since the firm could otherwise decrease m , and save on monitoring costs.

If there is no monitoring cost, i.e., if $C(m) = 0$ for any m , (C.2) will be identical to (C.1). The reason is quite simple. The firm can perfectly monitor workers without incurring any cost. That is, the optimal choice of m must be 1. Since workers' expected utility does not depend on m , nonshirking workers will not resist the firm's perfect monitoring. If $m = 1$, (C.2-2) becomes

$$U(w^e, e) \geq U(0, 0)$$

Obviously, this condition must hold even for the solution of (C.1), because otherwise no one will work. Also, the firm's profit function in (C.2) is exactly identical to that in (C.1). This means that when monitoring cost is arbitrarily

small, there is no significant difference between (C.1) and (C.2). Actually, (C.1) is a special case of (C.2), which can be obtained under the assumption of zero monitoring cost.

Assuming that the solution always satisfies (C.2-3), (C.2) predicts fixed wage compensation across states and layoffs in bad states.

PROPOSITION 2. The solution to (C.2) satisfies followings:

$$w^u(s) = \underline{w}^u$$

$$w^e(s) = \underline{w}^e$$

$$e(s) = \underline{e}$$

$$m(s) = \underline{m} \text{ for all } s, \text{ and}$$

$$dE/ds \geq 0$$

Proof. To solve (C.2), we can construct Lagrangean:

$$L = sf(eE) - w^e E - w^u (N-E) - C(m)E$$

$$+ \theta \{ (E/N)U(w^e, e) + (1-E/N)U(w^u, 0) \} + \phi G(w^e, e, m)$$

where $G(w^e, e, m) = U(w^e, e) - (1-m)U(w^e, 0) - mU(0, 0)$. Note that θ is independent of s , while ϕ is a function of s . From the first order condition, we have

$$(6) \quad U_w^u - N/\theta = 0$$

$$(7) \quad U_w^e(\theta/N) + \phi G_m/E - 1 = 0$$

$$(8) \quad sf' + U_e^e(\theta/N) + \phi G_e/E = 0$$

$$(9) \quad C' - \phi G_m/E = 0$$

$$(10) \quad sf'e - w^e + w^u - C + (U^e - U^u)(\theta/N) = 0$$

$$(11) \quad G = 0$$

where each subscript denotes the derivative with respect to the variable it represents, and $U^e = U(w^e, e)$, $U^u = U(w^u, 0)$. Since θ is independent of s , w^u is also independent of s . This means that w^u is a constant ($w^u = \underline{w}^u$). By substituting (9) into (7), (8), (10), and (11), we can rewrite the last five equations as follows:

$$(12) \quad U_w^e(\theta/N) + C'G_w/G_m - 1 = 0$$

$$(13) \quad sf' + U_e^e(\theta/N) + C'G_e/G_m = 0$$

$$(14) \quad C' - \phi G_m/E = 0$$

$$(15) \quad sf'e - w^e + w^u - C + (U^e - U^u)(\theta/N) = 0$$

$$(16) \quad G(w^e, e, m) = 0$$

Total differentiation and application of Cramer's rule of these equations yields⁴

$$dw^e/ds = de/ds = dm/ds = 0$$

$$dE/ds = -f'/f''_e > 0$$

QED

It is very hard to provide clear-cut explanation for these results, because all of the variables are interrelated in a complex manner. However, some partial intuition follows. The existence of monitoring costs gives the firm an incentive to decrease its level of employment. As we saw in (C.1), if there is no monitoring cost, the optimal contract is characterized by full employment in all states (to provide insurance to the risk-averse workers.) This is possible because work effort is perfectly substitutable for employment in production process. Workers are willing to accept lower wages in order to guarantee employment. However, monitoring

costs can be regarded as a fixed cost of employment. Therefore, in bad states, the firm would prefer lay-offs to save on the fixed cost of employment. Furthermore, since the employment level has no effect on NSC (see (C.2-2)), the firm has more discretion in choosing the employment level. This explains the employment fluctuation result. In this case, the firm must compensate workers with higher wages even in the bad states to make them bear the risk of being unemployed. This causes the wage profile across states to be flatter.

Another interesting result is that the optimal contract fixes the level of work effort. This is consistent with observation that unions usually try to predetermine worker's on-the-job duties in labor contracts.⁵ There is an conventional explanation about this phenomenon. If the firm has discretion on using workers, labor productivity could be increased, because firm could deploy its employees efficiently. Higher labor productivity will allow the firm to produce the given quantity of output with a lower level of employment. Therefore, the firm will have a smaller incentive to increase employment, if there is some fixed cost of employment. Therefore, unions resist increasing labor productivity. This interpretation is supported by my model. In bad states, the firm would prefer to increase the workers' level of effort with higher wages, decreasing total employment to save on monitoring cost. However, since risk-averse workers put higher value on employment than on wages, they will resist this strategy. Also, the firm's profitability of

adopting this strategy is limited. Higher work effort decreases nonshirking workers' utility. Higher wages increase shirkers' expected utility as well as nonshirkers' utility. Therefore, the firm has to increase monitoring intensity and cost. Different from the common explanation, (C.2) suggests that job descriptions are not only for workers' interest but also for that of firms.

III. Share Contract Model

Section II provided a model which explains lay-offs and fixed wage compensation during a contract period. In this section, I will show the Pareto-dominance of share contracts over wage contracts. The wage compensation per employee under a share contract can be defined as following:

$$(17) \quad w^e = v + \alpha sf(eE)/E$$

where v is a fixed component of compensation, and α is a share parameter that specifies the variable component of compensation. A share contract, δ_s is defined by

$$\delta_s = \{E(s), e(s), v(s), \alpha(s), m(s)\}$$

Begin by assuming that the contract agreement is enforced voluntarily, so that we may ignore the NSC. In this case, no share contract can Pareto-dominate the optimal wage contract. In fact, the optimal share contract is identical to that of the wage contract. The reason is quite simple. Under a share contract, wage compensation is decomposed into two parameters -- v and α . However, it is impossible to determine v and α separately, because the first order conditions for v and α are

identical. Without the NSC, the optimal share contract, δ_s^* , solves

$$(C.3) \max_{\delta} E_s \{ (1-\alpha) sf(e(s)E(s)) - v(s)E(s) - C(m(s))E(s) - w^u(s)(N-E(s)) \}$$

subject to

$$(C.3-1) E_s \{ (E(s)/N)U(\alpha sf(e(s)E(s))/E(s) + v(s), e(s)) + (1-E(s)/N)U(w^u(s), 0) \} \geq U_0$$

$$(C.3-2) 0 \leq E/N \leq 1$$

The first order conditions for v and α are given:

$$-E + (\theta/N)EU_w^e = 0$$

$$-sf(eE) + (\theta/N)sf(eE)U_w^e = 0$$

Both equations are reduced to

$$U_w^e = N/\theta$$

Therefore, we cannot determine v and α separately. Instead we can derive only an optimal combination of v and α . This result implies an interesting characteristic of the optimal share contract, as summarized in following proposition.

PROPOSITION 3. When workers have no incentive to shirk, the optimal share contract is identical to the optimal wage contract.

Proof. Define

$$w^e = \alpha sf(eE)/E + v$$

Substitute w^e into (C.3), and find the solution. This must be the solution to (C.1). From $w^e(s)$, we can obtain an optimal combination of v and α . QED

Note that the solution to (C.1) predicts full employment in all states. Basically, a contract between the firm and its workers provides insurance that cannot be obtained in market, due to the nontransferable characteristic of human capital. If both parties have perfect information about each other, the contract will be Pareto optimal. This result is nothing more than the optimal resource allocation under Debreu-Arrow's world of uncertainty. Therefore, no reform of the wage scheme could make both the firm and workers better off concurrently. However, insurance markets usually suffer from the *Moral Hazard* problem, caused by insurance companies' imperfect information on customers' behaviour. As we saw in section II, the workers' incentive to shirk leads to unemployment under the optimal wage contract. To prevent workers from shirking, the firm wastes its resources in monitoring workers. In this case, a wage compensation scheme which can suppress the incentive to shirk may improve the performance of the economy. This turns out to be true.

Consider the NSC in (C.2-2). Under a wage contract, shirkers do not suffer from their own shirking, since they still receive the same compensation as nonshirkers do. Under a share contract, however, shirkers do suffer from their own shirking. If a worker shirks, the total output actually produced will be smaller than the amount agreed to be produced. Workers get lower compensation, because some portion of wages is related to total revenue. The wage under a share contract is given by:

$$\alpha sf(e(E-S))/E + v$$

where S is the number of shirkers. When a worker decides to shirk, his expected utility is given:

$$(18) \quad (1-m)U(\alpha sf(e(E-1))/E+v, 0) + mU(0, 0)$$

Therefore, the NSC under a share contract can be expressed as

$$(C.3-3) \quad U(\alpha sf(Ee)/E+v, e) \geq (1-m)U(\alpha sf(e(E-1))/E+v, 0) + mU(0, 0)$$

Comparing (C.3-3) with (C.2-3), we can easily see that shirkers have lower expected utility under a share contract.

Hence, the firm could reduce m , and $C(m)$. This implies that firm and workers could be better off under a share contract.

This is stated in following proposition.

PROPOSITION 4. When workers have incentive to shirk, there exists a share contract which Pareto-dominates the optimal wage contract.

PROOF. The optimal wage contract is characterized by

$$\delta_w^* = \{\underline{w}^e, \underline{w}^u, \underline{e}, E^*(s), \underline{m}\}$$

Consider the following share contract,

$$\delta_s = \{\alpha(s), v(s), w^u(s), e(s), E(s), m(s)\}$$

with

$$E(s) = E^*(s),$$

$$e(s) = \underline{e},$$

$$\alpha(s) sf(e(s)E(s))/E(s) + v(s) = \underline{w}^e,$$

$$w^u(s) = \underline{w}^u, \quad \text{and}$$

$$m(s) = \underline{m} \quad \text{for all } s.$$

For all states, both contracts yield the same level of profit for the firm and the same level of expected utility for the workers. However,

$$U(\alpha f(\underline{e}E(s))/E(s)+v(s), 0) > U(\alpha f(\underline{e}(E(s)-1))/E(s)+v(s), 0)$$

Therefore,

$$\begin{aligned} U(\alpha f(\underline{e}E(s))/E(s)+v(s), e(s)) \\ &= U(\underline{w}^e, e(s)) \\ &= (1-\underline{m})U(\underline{w}^e, 0) + \underline{m}U(0, 0) \\ &= (1-\underline{m})U(\alpha f(\underline{e}E(s))/E(s)+v(s), 0) + \underline{m}U(0, 0) \\ &> (1-\underline{m})U(\alpha f(\underline{e}(E(s)-1))/E(s)+v(s), 0) + \underline{m}U(0, 0) \end{aligned}$$

The NSC is not binding under the share contract. Therefore, the firm can reduce \underline{m} , thereby decreasing $C(\underline{m})$, and increasing profits. QED

For any share contract, the bigger is the portion of compensation related to total revenue, the higher cost a shirker bears. Therefore, the best way to reduce the workers' incentive to shirk is to increase the share parameter, α , as much as possible. This gives us following result.

PROPOSITION 5. For any share contract, the optimal combination of α and v requires $v = 0$.

PROOF. Suppose not. Consider a share contract specifying $\delta_s = \{\alpha, v, e, E, \underline{m}\}$. I will suppress s for notational convenience. Suppose $v \neq 0$ at some s_0 . Choose $\underline{\alpha}$ such that $\underline{\alpha} f(Ee)/E = \alpha f(eE)/E + v$.

Consider a contract, δ , which is identical to δ_g except that \underline{q} replaces α . Then, firm's profits and workers' utility do not change under both contracts. However,

$$\begin{aligned}
 & \alpha f(e(E-1))/E + v \\
 &= \{\alpha f(eE)/E + v\} \{f(e(E-1))/f(eE)\} + v\{1-f(e(E-1))/f(eE)\} \\
 &= \{\underline{q}f(eE)/E\} \{f(e(E-1))/f(eE)\} + v\{1-f(e(E-1))/f(eE)\} \\
 &= \underline{q}f(e(E-1))/E + v\{1-f(e(E-1))/f(eE)\} \\
 &> \underline{q}f(e(E-1))/E
 \end{aligned}$$

Therefore,

$$U(\underline{q}f(eE)/E, e) > (1-m)U(\underline{q}f(e(E-1))/E, e) + mU(0, 0)$$

Firm can decrease m , thereby increasing profits at s_0 . This is a contradiction. QED

A share contract may cause (on average) more unemployment than a wage contract, even though all agents could be better off under that share contract. Higher unemployment in an economy will generate a contraction in aggregate demand, and thereby worsen all firms' economic positions. This implies a downward shift in the distribution of s , which firms confront. If this is true, that kind of share contract may not be desirable at a macroeconomic level. Therefore, my next question is whether a share contract could guarantee higher employment. The answer is affirmative as summarized in following proposition.

PROPOSITION 6. There exists a share contract which is Pareto-superior to, and has no less employment than the optimal wage contract specified in (C.2).

PROOF. The optimal wage contract is denoted by

$$\delta_w^* = \{\underline{w}^e, \underline{w}^u, \underline{e}, E^*(s), \underline{m}\}.$$

Consider a share contract, $\delta_s = \{\alpha(s), w^u(s), e(s), E(s), m(s)\}$

satisfying

$$(19) \quad e(s) = \underline{e},$$

$$(20) \quad \alpha(s) sf(\underline{e}E(s))/E(s) = \underline{w}^e,$$

$$(21) \quad w^u = \underline{w}^u,$$

$$(22) \quad U(\alpha(s) sf(e(s)E(s))/E(s), e(s))$$

$$= (1-m(s))U(\alpha(s) sf(e(s)(E(s)-1))/E(s), 0) + m(s)U(0, 0)$$

for all states. Note that $\alpha sf(\underline{e}(E-1))/E, 0 < \underline{w}^e$, for any E ,

as long as (20) holds. Consider the case in which $E(s) = E^*(s)$.

Since $U(\underline{w}^e, \underline{e}) = (1-\underline{m})U(\underline{w}^e, 0) + \underline{m}U(0, 0)$, $m(s) < \underline{m}$ for any s .

Therefore, δ_s must Pareto-dominate δ_w^* , when $E(s) = E^*(s)$.

Then the optimal form, δ_s^* , of δ_s 's also must Pareto-dominate

δ_w^* . This means that both the firm and the workers are better

off under δ_s^* . By Bellman's principle, for any state, both

are better off under δ_s^* . Let $\delta_s^* = \{\alpha^*(s), \underline{w}^u, \underline{e}, E^{**}(s),$

$m^*(s)\}$. Suppose that $E^{**}(s) < E^*(s)$, for some s_0 . Then,

$$(E^{**}/N)U(\alpha^* sf(\underline{e}E^{**})/E^{**}, \underline{e}) + (1-E^{**}/N)U(\underline{w}^u, 0)$$

$$= (E^{**}/N)U(\underline{w}^e, \underline{e}) + (1-E^{**}/N)U(\underline{w}^u, 0)$$

$$< (E^*/N)U(\underline{w}^e, \underline{e}) + (1-E^*/N)U(\underline{w}^u, 0)$$

This shows that workers have lower expected utility under δ_s^* at s_0 . This is a contradiction. QED

Now, in addition to Pareto dominating wage contracts, a share contract generates less employment fluctuation than the optimal wage contract. If workers are more likely to be employed, there will be an increase in aggregate consumption, which shifts up distributions of s , and improves firms' profitability. An interesting implication of Proposition 6 is that this is possible even when wages actually given to workers are constant across states. Therefore, share contracts will Pareto-dominate even at the macro level. One shortcoming of Weitzman's analysis is that a firm has excess demand for labor only if workers already employed are willing to accept lower wages. (See Nordhaus [1988].) My model avoids this problem.

IV. Conclusion

In contrast to Weitzman's work, my model, which is based on a framework of implicit contract theory, allows us to make a complete welfare comparison between two different wage compensation schemes -- fixed-wage and share compensations. Fixed wages across the states are predicted rather than assumed in an *ad hoc* way. Therefore, the welfare comparison given this paper is less open to criticism than comparisons given by other studies.

Wage contracts can be regarded as a special form of share contracts. A share contract predetermines some fixed portion of wage compensation, with the remaining portion tied to the firm's performance in the product market. A fixed-wage contract, which is the optimal wage contract in my model, is a share contract with no variable portion of compensation. This suggests that fixed-wage contracts would be a suboptimal choice among share contracts. My model shows that this is the case. If workers have no incentive to shirk on their job, the wage and the optimal share contracts are identical. Wage contracts are always characterized by full employment in order to insure risk-averse workers. However, if workers have some incentive to shirk, the optimal wage contract specifies fixed wages across the states of nature, generating lay-offs in bad states. In this case, there exists a share contract which not only Pareto-dominates the optimal wage contract, but also has no less employment for any state. This is possible because workers' incentive to shirk decreases under share contracts.

In Weitzman's model, firms could have excess demand for labor only when their employed workers are willing to accept lower wages. The employed workers will accept only if the share contracts they accept generate sufficient *macroeconomic externalities* on the whole economy that their suspended self-interests are ultimately compensated. For this to be true, substantially large portion of sectors in the economy must adopt the share contracts, and the share parameters should be based on the exact information on the economy, all of which

seem to be practically difficult. However, in my model, no one suffers from a share contract even at a micro level. No one has to wait until the *macroeconomic externalities* compensate his suspended self-interest. Furthermore, the share contracts will increase the economy-wide employment level, even when the share sectors are small. Exact information on the economy is not required to find this form of share contracts, as suggested by PROPOSITION 6. Therefore, the implementation of the share contracts will not be very costly.

However, even though my model provides an argument in favor of the share economy, it is early to draw some policy conclusions. Unfortunately, my model fails to provide a clear-cut answer to a different question: Why does an economy resist converting from a wage to a share economy, if share contracts are really superior to wage contracts? I can offer only some partial intuition. Under a share contract, each worker's wage depends on other agents' work effort. If there is a shirker, all workers suffer from lower wages. Therefore, some burden of monitoring should fall on the workers themselves -- each worker becomes more sensitive to the other workers' behavior. This generates some kind of psychological cost to workers. In this situation, workers may prefer wage contracts. From standpoint of the firms, a share contract will reduce a management's discretion on production. A share contract can be successful, only if workers and a firm share the exact information concerning the firm's real market

situation and its true total revenue. In other words, a credibility problem arises in a share economy. This means labor's participation in management is a necessary condition for a successful share contract. In this case, a management cannot efficiently cope with an abrupt change in the firm's market situation, because any decision of the management should wait for its workers' approval. A longer decision process will reduce the firm's profitability. Therefore, firms might also prefer wage contracts. These are just some possible reasons why an economy might be characterized by wage contracts. Clearly, this question requires further study.

ENDNOTES

1. Weitzman [1985], p. 3.
2. Ibid., p. 3.
3. Ibid., pp 118-122.
4. The total differentiation of (13), (14), (15), (16), and (17) yields:

$$\begin{bmatrix}
 0 & U_{ww}^e(\theta/N) + C'G_{ww}/G_m - C'G_w G_{mw}/G_m^2 & U_{we}^e(\theta/N) + C'G_{we}/G_m \\
 0 & U_{ew}^e(\theta/N) + C'G_{ew}/G_m - C'G_w G_{mw}/G_m^2 & sf''E + U_{ee}^e(\theta/N) + C'G_{ee}/G_m \\
 -G_m/E & -\phi G_{mw}/E & 0 \\
 0 & U_w^e(\theta/N) - 1 & sf''eE + sf' + U_e^e(\theta/N) \\
 0 & G_w & G_e
 \end{bmatrix}$$

$$\begin{bmatrix}
 C''G_w/G_m + C'G_{wm}/G_m & 0 \\
 C''G_e/G_m & sf''e \\
 C'' & \phi G_m/E^2 \\
 -C' & sf''e^2 \\
 G_m & 0
 \end{bmatrix}
 \begin{bmatrix}
 d\phi \\
 dw_e \\
 de \\
 dE \\
 dm
 \end{bmatrix}
 =
 \begin{bmatrix}
 0 \\
 -f' \\
 0 \\
 -f'e \\
 0
 \end{bmatrix}
 ds$$

5. See Balfour [1987], pp 300-328.

APPENDIX

APPENDIX

In Section II, I assumed that each worker's hours on the job are fixed. This assumption is not required to obtain fixed wage under the optimal wage contract. To show this, redefine e as the hourly effort level of a worker, and w^e as the hourly wage rate. I assume that working hours (h) are perfect substitute for hourly working effort (e) in the production process. First, suppose that workers have no incentive to shirk on the job. Let

$$\delta_w = \{E, e, h, w^e, w^u\}$$

For simplicity, I suppress s . The optimal contract, δ_w^* , solves

$$(A.1) \max_{\delta} E_s \pi = E_s \{sf(ehE) - w^e h E - w^u (N - E)\}$$

subject to

$$(A.1-1) \quad E_s \{(E/N)U(w^e, h, eh) + (1 - E/N)U(w^u, 0)\} \geq U_0$$

We can easily show that PROPOSITION 1 still holds for (A.1). I omit the proof, since it is similar to that of PROPOSITION 1.

Now, consider the case in which workers have incentive to shirk. Describe a wage contract by

$$\delta_w = \{E, e, h, w^e, w^u, m\}.$$

The optimal contract δ_w^* , solves

$$(A.2) \max_{\delta} E_s \{sf(ehE) - w^e h E - w^u (N - E) - C(m)E\}$$

subject to

$$(A.2-1) \quad E_s \{(E/N)U(w^e h, eh) + (1 - E/N)U(w^u, 0)\} \geq U_0$$

$$(A.2-2) \quad U(w^e h, eh) \geq (1 - m)U(w^e h, 0) + mU(0, 0)$$

Consider the first order conditions with respect to e , w^e , and h :

$$(1) \quad sf'hE + \theta(E/N)U_e^e h + \phi U_e^e h = 0$$

$$(2) \quad -E + (E/N)U_w^e + \phi(U_w^e - (1-m)U_w^s) = 0$$

$$(3) \quad sf'eE - w^e E + \theta((E/N)U_w^e w^e + (E/N)U_e^e e) \\ + \phi(U_w^e w^e + U_e^e e - (1-m)U_w^s w^s) = 0$$

where U^s denotes the utility of shirkers, $U(w^e h, 0)$. Observe that one of (1), (2), and (3) are a linear combination of the others. This means that we cannot determine e , w^e , and h , separately. To avoid this problem, let

$$e^o = eh$$

$$w = w^e h.$$

Substitution of e^o and w into (A.2) gives us (C.2). Therefore, the optimal choice, across the states of nature, must fix e^o and w at \underline{e}^o and \underline{w} , respectively. Total wage compensation for each worker (w) and total working effort per worker (e^o) are independent of the states. Firm can arbitrarily choose fixed working hours before a contract. Then, as in PROPOSITION 2, we have fixed hourly wage and effort level under the optimal contract.

(A.2) contains another interesting implication. One may consider the case in which firm predetermines e at some level regardless of the states. Suppose that $e=1$. Then, (A.2) is still identical to (C.2) except that h replaces e , and that w ($=w^e h$) takes the role of w^e in (C.2). Substituting w and h into and solving (C.2) will generate constant w and h . That is, when e is predetermined, the optimal contract has the

constant w^e during a contract period. This result is surprising, because conventional implicit contract models usually fail to predict fixed wage rates when working hours are allowed to vary. If effort level is fixed, and if NSC (A.2-2) is ignored, (A.2) becomes a standard contract model. (See p. 8 in Cooper [1987].) Without NSC, (A.2) fails to generate fixed wage rate, unless there are some fairly strong assumptions on the form of workers' utility function and on variations in hours. Therefore, NSC in (A.2) has a crucial role in generating fixed wages across the states of nature under the optimal contract. By this reasoning, it is safe to say that we do not have to assume fixed working hours.

REFERENCES

REFERENCES

- Balfour, A. (1987), Union-Management Relations in a Changing Economy, Prentice-Hall, Inc., Englewood Cliff.
- Cooper, R. (1987), Wage and Employment Patterns in Labor Contracts: Microfoundations and Macroeconomic Implications, Harwood Academic Publishers, New York.
- Cooper, R. (1988), "Will Share Contracts Increase Economic Welfare?," The American Economic Review, Vol 78, pp 139 - 154.
- Cooper, R. and John A. (1988), "Coordinating Coordination Failure in the Keynesian Models," Quarterly Journal of Economics, August, pp 441 - 463.
- John, A. (1987), "Employment Fluctuations in a "Share Economy", Working Paper, Michigan State University.
- Miller, R.E. (1979), Dynamic Optimization and Economic Applications, McGraw Hill Inc.
- Milton, H. and Artur, R. (1978), "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement," The American Economic Review, Vol. 68, pp 20 -30.
- Nordhaus, W. (1988), "Can the Share Economy Conquer Stagflation?" Quarterly Journal of Economics, February, pp 201 - 223.
- Nordhaus, W. and John, A., eds. (1986), "The Share Economy: A Symposium," Journal of Comparative Economics, Vol. 10, pp 414 - 473.
- Shapiro, C. and Stiglitz, J.E. (1984), "Equilibrium Unemployment as a Worker Discipline Device," The American Economic Review, Vol. 74, pp 433 - 444.
- Weitzman, M.L. (1984), The Share Economy, Harvard University Press, Cambridge.
- Weitzman, M.L. (1984), "Profit Sharing as Macroeconomic Policy," AEA Papers and Proceedings, Vol. 75, pp 41 - 45.
- Weitzman, M.L. (1985), "The Simple Macroeconomics of Profit Sharing," The American Economic Review, Vol. 75, pp 937 - 953.
- Weitzman, M.L. (1987), "Steady State Unemployment Under Profit Sharing," The Economic Journal, Vol. 97, pp 86 - 105.

Chapter 2

The Joint Estimation of a Model of Labor Force and Employment Decisions and Market Wages

I. Introduction

Much of the empirical literature on the labor supply decision simply assumes that individuals can obtain jobs once they decide to enter the labor market. That is, the employment status of an individual is determined by only one selection criterion -- the individual's decision on whether to work. In many models used to explain employment status, individuals are categorized into two groups: employed and nonemployed. The unemployed and non-participants are treated as behaviorally identical in their decision process, and both are regarded as one single group -- the nonemployed. These models implicitly ignore frictions in the labor market. It is a well-known fact that unemployment is not simply explained by individuals' work incentives. According to the single selectivity criterion based on employment and nonemployment, both the unemployed and people not in the labor force (NLF) choose not to work because their reservation wage rates are greater than their market wage rates. Therefore, the unemployment status of an individual is purely voluntary. In this sense, I call these traditional models No-Friction models. The aim of this paper is to provide estimates of the determinants of the employment status of married women and their market wages by using two different selection criteria

-- preferences for work and ability to become employed.

Some studies show the importance of the existence of the unemployed in a given data set. Flinn and Heckman [1983], applying a duration model to young men selected from the National Longitudinal Survey, reject the hypothesis that the classifications unemployed and NLF are behaviorally equivalent. Ham [1982], using a sample of prime aged males taken from the University of Michigan's Panel Study of Income Dynamics (PSID), shows that the estimates of parameters in an equation for work hours are biased if the unemployed or underemployed workers are ignored. Also, Blundell, Ham, and Meghir [1987] reject the Tobit model based on the traditional No-Friction model, using a sample of married women drawn from the UK Family Expenditure Survey of 1981.

We categorize individuals into three different groups: employed, unemployed, and non-participants. A married woman is assumed to enter the labor market if her reservation wage is less than the prevailing market wage. However, not all individuals who decide to enter the labor market get jobs immediately. A woman is employed only when she matches with an employer who is willing to hire her. An individual who is better able to find potential employers will have a higher probability of being employed. In this sense, I call the model in this paper a Friction model. For this model, we may construct a job-match equation which can distinguish between the employed and the unemployed in a probit framework. Labor-force and employment decisions can be jointly explained by a

bivariate probit model with partial observability. (See Meng and Schmidt [1985], and Farber [1983].)

This paper also estimates the parameters in the wage equation by the maximum-likelihood estimation (MLE) method. We observe only the currently employed workers' wages. That is, the observed distribution of wages depends not only on individuals' decisions about labor-force participation, but also on their ability to find jobs. The data collected, therefore, will have two types of selection biases. If we estimate labor-force and employment decisions and the wage equation jointly, the parameters in equations for labor-force and employment decisions will be estimated more efficiently. At the same time, the conventional loglikelihood ratio (LR) test is applicable for the hypothesis of no selection bias.

The extension of Heckman's simple two-stage estimation method is usually used in other studies for cases where two selection rules generate the sample. (See Fische, Trost, and Lurie [1981], and Ham [1982].) The selectivity regressors used in the least squares (OLS) estimation of the wage equation are generated by a bivariate probit model. We can easily apply this extended two-stage estimation method to the case where employed, unemployed, and NLF people are observed separately. Other studies usually use an F-statistic for the test of the joint significance of the selectivity regressors. In Heckman's simple selectivity model, the standard t-statistic for the selectivity regressor has been used for the test of no selection bias. Melino[1982] (also, Lin[1982])

shows that this t-statistic is asymptotically equivalent to the Lagrangean Multiplier (LM) test statistic. Likewise, this paper shows that the F statistic for the hypothesis of no selection bias for the model with two selection rules is asymptotically equivalent to the LM test. This means that the F-statistic has good power properties, at least asymptotically.

The empirical results in this paper reveal that the Friction model explains a married woman's labor status better than the No-Friction model. The joint estimation of labor-force, employment status and the wage rate generates more reliable estimates. There is significant evidence of selection bias from ignoring labor-force and employment status. The correction for sample selection bias in the extended two-stage method turns out to be not quite as satisfactory as that by the MLE method.

This paper is organized in the following way. Section II describes the basic model for frictions in the labor market. Section III summarizes the data, and describes the explanatory variables used for the empirical study. Section IV demonstrates the empirical results. Some concluding remarks follow in Section V.

II. Model

This section explains the basic model based on the assumption of frictions in the labor market. Notice that not all married women who enter the labor market get jobs. We may

assume that if an individual has higher market wage than her reservation wage, she enters the labor market. However, unless she matches an employer who is willing to hire her, she remains unemployed. The Current Population Survey (CPS), the National Longitudinal Survey (NLS), and the Panel Study of Income Dynamics (PSID) provide the data on three different groups of married women: the employed, the unemployed, and the non-participants. In this case, we need another employment criterion to consider the unemployed people separately from the NLF people. We may imagine that each individual has her own job-match skill. Those who have better job match ability will have a higher probability of being employed. Let M_i^* be the index for the i 'th individual's job-match skill; w_i , the market wage; w_i^R , the reservation wage. If $w_i \geq w_i^R$, the i 'th individual participates in the labor market. If $M_i^* \geq 0$, she matches an employer, and get a job which pays her w_i . If $w_i < w_i^R$, she retains her NLF status. When $w_i \geq w_i^*$ and $M_i^* < 0$, she is in the labor market, and remains unemployed. w_i is observed if and only if the i 'th individual is employed. Therefore, individuals' behavior in the labor market is determined by the three variables -- market wage rates, reservation wage rates, and the job-match index. These are summarized by Model I.

Model I. The Structural Model

$$(1.1) \quad w_i^R = z_{1i}\delta_1 + \epsilon_{1i}$$

$$(1.2) \quad M_i^* = z_{2i}\delta_2 + \epsilon_{2i}$$

$$(1.3) \quad w_i = z_{3i}\delta_3 + \epsilon_{3i}$$

$$\begin{bmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & 1 & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} \end{bmatrix} \right], \quad i = 1, 2, \dots, N.$$

$(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})'$ are independently and identically distributed. The z_{ji} are the observed $1 \times k_j$ vectors of explanatory variables.

$$y_{1i} = \begin{cases} 1, & \text{if } w_i \geq w_i^R \\ 0, & \text{otherwise.} \end{cases}$$

$$y_{2i} = \begin{cases} 1, & \text{if } M_i^* \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

y_{2i} is observed if and only if $y_{1i}=1$. w_i^R and M_i^* are not observed. w_i is observed if and only if $y_{1i}y_{2i}=1$. y_{1i} denotes labor-force status (LF), and y_{2i} denotes ability to find a job. Therefore an individual is employed if $y_{1i}y_{2i}=1$.

For a simple estimation procedure for this model, let

$$(2.1) \quad e_{1i} = (\epsilon_{3i} - \epsilon_{1i}) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2}$$

$$(2.2) \quad e_{2i} = \epsilon_{2i}$$

$$(2.3) \quad e_{3i} = \epsilon_{3i}$$

Then,

$$(e_{1i}, e_{2i}, e_{3i})' \sim N(0, \Omega)$$

where

$$\Omega = \begin{bmatrix} 1 & \rho & \sigma_{13} \\ \cdot & 1 & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & (\Sigma_{23} - \Sigma_{12}) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2} & (\Sigma_{33} - \Sigma_{13}) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2} \\ \cdot & 1 & \Sigma_{23} \\ \cdot & \cdot & \Sigma_{33} \end{bmatrix}$$

Let

$$\begin{aligned} (3.1) \quad y_{1i}^* &= (w_i - w_i^R) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2} \\ &= (z_{3i}\delta_3 - z_{1i}\delta_1) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2} + e_{1i} \\ &= x_{1i}\delta_1 + e_{1i} \end{aligned}$$

$$(3.2) \quad y_{2i}^* = M_i^* = z_{2i}\delta_2 + \epsilon_{2i} = x_{2i}\delta_2 + e_{2i}$$

$$(3.3) \quad y_{3i} = w_i = z_{3i}\delta_3 + \epsilon_{1i} = x_{3i}\delta_3 + e_{3i}$$

Now, we can rewrite Model I as follows.

Model II. Reduced-Form Model.

$$(4.1) \quad y_{1i}^* = x_{1i}\beta_1 + e_{1i}$$

$$(4.2) \quad y_{2i}^* = x_{2i}\beta_2 + e_{2i}$$

$$(4.3) \quad y_{3i} = x_{3i}\beta_3 + e_{3i}$$

$$\begin{bmatrix} e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix} \sim N(0, \Omega) \quad i = 1, 2, \dots, N.$$

where

$$\Omega = \begin{bmatrix} 1 & \rho & \sigma_{13} \\ \cdot & 1 & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{bmatrix}$$

$$y_{1i} = \begin{cases} 1, & \text{if } y_{1i}^* \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$y_{2i} = \begin{cases} 1, & \text{if } y_{2i}^* \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

y_{2i} is observed if and only if $y_{1i}=1$. y_{1i}^* and y_{2i}^* are not observed; y_{3i} is observed if and only if $y_{1i}y_{2i}=1$.

If our interest is just in the joint estimation of labor-force and employment decisions, the parameters in those processes can be estimated by a bivariate probit method with partial observability. (Specifically, this is the "Censored probit" model of Farber [1983] and Meng and Schmidt [1985].) Here we may distinguish two cases, depending on what is observed. First, there is the case of partial observability in the sense of Poirier [1980], in which we do not observe y_{1i} or y_{2i} for anyone, but we observe $(y_{1i}y_{2i})$. This corresponds to observing only employment status (employed or not), but not labor-force participation status for individuals who are not employed. Interestingly, we can still estimate separately the labor-force participation and employment equations, using Poirier's model.

Case I. Partial Observability in the sense of Poirier

$$y_{1i} = \begin{cases} 1, & \text{if } y_{1i}^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$y_{2i} = \begin{cases} 1, & \text{if } y_{1i}^* \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad i = 1, 2, \dots, N.$$

Only $y_{1i}y_{2i}$ is observed. y_{1i} is one if the i 'th person in the labor-force, and y_{2i} is one if the i 'th person is able to find a job. Thus $y_{1i}y_{2i}$ is one if the i 'th person is employed.

In this case, the maximum likelihood (ML) estimators of β_1 , β_2 , and ρ are derived by maximizing the following log-likelihood function with respect to β_1 , β_2 , and ρ :

$$(5) \quad \ln L^P(\beta_1, \beta_2, \rho) = \sum_{i=1}^N \{ y_{1i} y_{2i} \ln[F(x_{1i} \beta_1, x_{2i} \beta_2, \rho)] + (1 - y_{1i} y_{2i}) \ln[1 - F(x_{1i} \beta_1, x_{2i} \beta_2, \rho)] \}$$

where $F(\cdot)$ is the bivariate standard normal distribution function. This method can allow for frictions in the labor market. However, the problem with this model is that the information about who is unemployed is wasted. This model basically categorizes the individuals in a given data set into just two groups; the employed and the nonemployed. The latter group consists of those who are not willing to work (NLF people) and those who want to work but can not find jobs (unemployed). This model does not identify who are who among the nonemployed. Therefore, even though the estimates from this model would be consistent, these estimates are generally inefficient, if information about who is unemployed is available. (See Meng and Schmidt [1985].)

The second case we consider is the case in which we observe y_{2i} when $y_{1i}=1$ (though not when $y_{1i}=0$). Thus we can observe an individual's success or failure in job search only when she is in the labor market. The unemployed people are identified as those who are willing to work (being in the labor market) but are not successful in their job search. Since this model can distinguish the unemployed from the NLF people, we can get more efficient estimates than we would get from Poirier's model. In this case, the parameters can be

estimated by censored probit method.

Case II. Censored probit.

$$y_{1i} = \begin{cases} 1, & \text{if } y_{1i}^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$y_{2i} = \begin{cases} 1, & \text{if } y_{2i}^* \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

y_{1i} is observed. y_{2i} is observed if and only if $y_{1i} = 1$.

In this case, the ML estimators of β_1 , β_2 , and ρ are derived by maximizing the following log-likelihood function with respect to β_1 , β_2 , and ρ :

$$\begin{aligned} (6) \ln L^C(\beta_1, \beta_2, \rho) = & \sum_{i=1}^N \{ y_{1i} y_{2i} \ln[F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)] \\ & + y_{1i}(1-y_{2i}) \ln[\Phi(x_{1i}\beta_1) - F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)] \\ & + (1-y_{1i}) \ln(1-\Phi(x_{1i}\beta_1)) \} \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function. This paper uses the censored probit model to estimate LF and employment decisions.

The extension of Heckman's simple two-stage method is usually used when two different types of selection biases exist in the data set. The wage equation given in Model II also could be estimated by the extended two-stage method. The conditional mean of e_{3i} will be given by

$$\begin{aligned} (7) E(e_{3i} | y_{1i}^* \geq 0, y_{2i}^* \geq 0) = & E(e_{3i} | e_{1i} \geq -x_{1i}\beta_1, e_{2i} \geq -x_{2i}\beta_2) \\ & + \sigma_{13} \frac{\phi(x_{1i}\beta_1) \Phi[(x_{2i}\beta_2 - \rho x_{1i}\beta_1) / (1-\rho^2)^{1/2}]}{F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)} \end{aligned}$$

$$+ \sigma_{23} \frac{\phi(x_{2i}\beta_2) \Phi[(x_{1i}\beta_1 - \rho x_{2i}\beta_2) / (1-\rho^2)^{1/2}]}{F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)}$$

where

$$f(t_1, t_2, \rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2(1-\rho^2)} (t_1^2 - 2\rho t_1 t_2 + t_2^2)\right]$$

and,

$$F(h, k, \rho) = \int_{-\infty}^h \int_{-\infty}^k f(t_1, t_2, \rho) dt_2 dt_1.$$

where $\phi(\cdot)$ is the standard normal distribution function. We can find this formula in Fische, Trost, and Lurie [1981], Ham [1981], Maddala [1987], and Poirier [1980]. (Also, see APPENDIX A.) Now, we can rewrite (4.3) as

$$(4.3') \quad y_{3i} = x_{3i}\beta_3 + \sigma_{13}\mu_{1i} + \sigma_{23}\mu_{2i} + v_{3i}$$

where $E(v_{3i} | y_{1i}^* \geq 0, y_{2i}^* \geq 0) = 0$, and

$$\mu_{1i} = \frac{\phi(x_{1i}\beta_1) \Phi[(x_{2i}\beta_2 - \rho x_{1i}\beta_1) / (1-\rho^2)^{1/2}]}{F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)}$$

$$\mu_{2i} = \frac{\phi(x_{2i}\beta_2) \Phi[(x_{1i}\beta_1 - \rho x_{2i}\beta_2) / (1-\rho^2)^{1/2}]}{F(x_{1i}\beta_1, x_{2i}\beta_1, \rho)}$$

The extended two-stage estimator can be obtained by the following steps. First, estimate β_1 , β_2 , and ρ by the censored probit model, and estimate μ_{1i} and μ_{2i} ($\hat{\mu}_{1i}, \hat{\mu}_{2i}$) by using the estimated β_1 , β_2 , and ρ . We then have the estimates of β_1 , β_2 , ρ , μ_{1i} , μ_{2i} ; $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\rho}$, $\hat{\mu}_{1i}$, and $\hat{\mu}_{2i}$. Second, regress y_{3i} on x_{3i} , $\hat{\mu}_{1i}$, and $\hat{\mu}_{2i}$ by OLS, using only the observed y_{3i} . In order to find the asymptotic distribution of the extended two-stage estimators, we just assume for now that

$$(8) \sqrt{N}(\hat{r} - r) = \sqrt{N} \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \hat{\rho} - \rho \end{bmatrix} \rightarrow N(0, \Psi)$$

where $\hat{\beta}_1, \hat{\beta}_2, \hat{\rho}$ are the ML estimates of β_1, β_2 , and ρ generated by the censored probit model. Let N_1 denote the number of the observed y_{3i} . For simplicity, assume that $\text{plim } N_1/N = k$, $0 < k < 1$. Then, we can show (See APPENDIX B.) that

$$(9) \sqrt{N} \begin{bmatrix} \hat{\beta}_3 - \beta_3 \\ \hat{c} - c \end{bmatrix} \rightarrow N(0, DBD')$$

where

$$\mu_i = (\mu_{1i} \mu_{2i})$$

$$c = (\sigma_{13} \sigma_{23})$$

$$A_{1i} = \partial \mu_{1i} / \partial r' = [(-x_{1i}\beta_1\mu_{1i} - \mu_{1i}^2 - \rho\mu_{3i})x_{1i}, (\mu_{3i} - \mu_{1i}\mu_{2i})x_{2i}, \\ -((x_{1i}\beta_1 - \rho x_{2i}\beta_2) / (1 - \rho^2))\mu_{3i} - \mu_{1i}\mu_{3i}]$$

$$A_{2i} = \partial \mu_{2i} / \partial r' = [(\mu_{3i} - \mu_{1i}\mu_{2i})x_{1i}, (-x_{2i}\beta_2\mu_{2i} - \mu_{2i}^2 - \rho\mu_{3i})x_{2i}, \\ -((x_{2i}\beta_2 - \rho x_{1i}\beta_1) / (1 - \rho^2))\mu_{3i} - \mu_{2i}\mu_{3i}]$$

$$A_i = \partial \mu_i / \partial r' = [A_{1i}' \ A_{2i}']'$$

$$Q_{ij} = c'A_i \Psi A_j' c$$

$$\pi_{ii} = 1 - (\sigma_{13}^2 / \sigma_{33})(x_{1i}\beta_1\mu_{1i} + \mu_{1i}^2 + \rho\mu_{3i}) \\ - (\sigma_{23}^2 / \sigma_{33})(x_{2i}\beta_2\mu_{2i} + \mu_{2i}^2 + \rho\mu_{3i}) \\ + 2\sigma_{13}\sigma_{23}(\mu_{3i} - \mu_{1i}\mu_{2i}) / \sigma_{33}$$

$$D = \text{plim} \left[\frac{1}{N_1} \begin{bmatrix} \sum_i x_{3i}' x_{3i} & \sum_i x_{3i}' \mu_i \\ \sum_i \mu_i' x_{3i} & \sum_i \mu_i' \mu_i \end{bmatrix} \right]^{-1}$$

$$B = \sigma_{33} \cdot \text{plim} \frac{1}{N_1} \begin{bmatrix} \sum_i \pi_{ii} x_{3i}' x_{3i} & \sum_i \pi_{ii} x_{3i}' \mu_i \\ \sum_i \pi_{ii} \mu_i' x_{3i} & \sum_i \pi_{ii} \mu_i' \mu_i \end{bmatrix}$$

$$+ \text{plim} \frac{N_1}{N} \frac{1}{N_1^2} \begin{bmatrix} \sum_i \sum_j Q_{ij} x_{3i}' x_{3j} & \sum_i \sum_j Q_{ij} x_{3i}' \mu_i \\ \sum_i \sum_j Q_{ij} \mu_i' x_{3j} & \sum_i \sum_j Q_{ij} \mu_i' \mu_i \end{bmatrix}$$

where \sum_i denotes the summation over i from 1 to N_1 . If $\sigma_{13} = \sigma_{23} = 0$, then, $\pi_{ii} = 1$, and $Q_{ij} = 0$. In this case,

$$DBD' = \sigma_{33} \cdot \text{plim} \frac{1}{N_1} \begin{bmatrix} \sum_i x_{3i}' x_{3i} & \sum_i x_{3i}' \mu_i \\ \sum_i \mu_i' x_{3i} & \sum_i \mu_i' \mu_i \end{bmatrix}$$

which is the standard covariance matrix for the OLS estimator.

The more efficient are the estimators of β_1 , β_2 , and ρ that are used, the more efficient the estimators of β_3 and c will be. Suppose r_1 and r_2 are consistent estimators of r with corresponding asymptotic covariance matrices, Ψ_1 , and Ψ_2 , respectively. Consider the second matrix in B:

$$\begin{bmatrix} \sum_i \sum_j Q_{ij} x_{3i}' x_{3j} & \sum_i \sum_j Q_{ij} x_{3i}' \mu_i \\ \sum_i \sum_j Q_{ij} \mu_i' x_{3j} & \sum_i \sum_j Q_{ij} \mu_i' \mu_i \end{bmatrix} \\ = \begin{bmatrix} \sum_i x_{3i}' r' A_i \\ \sum_i \mu_i' r' A_i \end{bmatrix} \Psi \begin{bmatrix} \sum_i x_{3i}' r' A_i \\ \sum_i \mu_i' r' A_i \end{bmatrix}' = G \Psi G'$$

If $\Psi_1 - \Psi_2$ is positive semidefinite, so is $G \Psi_1 G' - G \Psi_2 G'$. This means that we can get more efficient estimators of β_3 and c using r_2 than r_1 . Meng and Schmidt [1985] provide the asymptotic covariance matrices of the estimates of β_1 , β_2 , and ρ for Case I and Case II. Their comparison of asymptotic efficiencies shows that Case II generates more efficient estimators of β_1 , β_2 , and ρ than Case I. This means that Case II is more desirable for the extended two-stage estimation than Case I.

This extended two-stage method does not give us estimates of all the parameters. We still have to estimate σ_{33} . For

th

ob

es

1

w

P

e

f

s

c

this, denote the residuals for the first N_1 observations obtained from the second stage as \hat{v}_{3i} , $i=1,2,\dots,N_1$. Then an estimate of σ_{33} is

$$10) \hat{\sigma}_{33} = (1/N_1) \sum_i \{ \hat{v}_{3i}^2 + \hat{\sigma}_{13}^2 [(x_{1i}\hat{\beta}_1)\hat{\mu}_{1i} + \hat{\mu}_{1i}^2 + \hat{\rho}\hat{\mu}_{3i}] \\ + \hat{\sigma}_{23}^2 [(x_{2i}\hat{\beta}_2)\hat{\mu}_{2i} + \hat{\mu}_{2i}^2 + \hat{\rho}\hat{\mu}_{3i}] - 2\hat{\sigma}_{13}\hat{\sigma}_{23}(\hat{\mu}_{3i} - \hat{\mu}_{1i}\hat{\mu}_{2i}) \}$$

where $\mu_{3i} = f(x_{1i}\beta_1, x_{2i}\beta_2, \rho) / F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)$. We can show that $\text{plim } \hat{\sigma}_{33} = \sigma_{33}$ (See APPENDIX A.) This completes the estimation of all parameters by the extended two-stage method for Model II.

In Heckman's simple selectivity model, the test for sample selectivity is usually done by the t-statistic for the estimated coefficient of the selectivity regressor. Melino [1982] and Lin [1982] show that this t-test is the same as the LM test of the hypothesis of no selectivity bias, so it has good asymptotic power properties. In the same way, we can perform an F-test for sample selectivity in Model II. If $\sigma_{13} = \sigma_{23} = 0$, then

$$E(e_{3i} | e_{1i} \geq -x_{1i}\beta_1, e_{2i} \geq -x_{2i}\beta_2) = 0.$$

Therefore, sample selection bias is not produced by applying OLS to (4.3) directly. Furthermore, as we see later, the conditional distribution of e_{3i} given $e_{1i} \geq -x_{1i}\beta_1$, and $e_{2i} \geq -x_{2i}\beta_2$ is normal and homoscedastic for all $i = 1, 2, \dots, N_1$. Hence, we can use the conventional F-statistic to test for sample selectivity. For notational convenience, let

$$X_3 = [x_{31}' \ x_{32}' \ \dots \ x_{3N_1}']'$$

$$\hat{\mu} = [\hat{\mu}_1' \ \hat{\mu}_2' \ \dots \ \hat{\mu}_{N_1}']'$$

$$\hat{Z} = [X_3 \quad \hat{\mu}]$$

$$\hat{\theta} = [\hat{\beta}_3' \quad \hat{c}']'$$

$$Y_3 = [Y_{31} \ Y_{32} \ \dots \ Y_{3N_1}]'$$

$$SSE_0 = (Y_3 - \hat{Z}\hat{\theta})'(Y_3 - \hat{Z}\hat{\theta})$$

$$M_X = I - X_3(X_3'X_3)^{-1}X_3'$$

Then, under the null hypothesis that $c=0$, the test statistic,

$$(11) \ F = \frac{\hat{c}'\hat{\mu}'M_X\hat{\mu}\hat{c}/2}{SSE_0/(N_1-k_3)}$$

is treated as if it were distributed as $F(2, N_1-k_3)$. It can be shown that this F-test is asymptotically identical to the LM test. (See APPENDIX C.)

Model II allows us to estimate equations for labor-force and employment decisions and the wage rate jointly. The log-likelihood function for Model II can be derived, after some straightforward but messy algebra. The conditional distribution of the e_{3i} is given, for $i = 1, 2, \dots, N_1$, as

$$(12) \ g(e_{3i} | e_{1i} \geq -x_{1i}\beta_1, e_{2i} \geq -x_{2i}\beta_2) \\ = \frac{1}{(2\pi)^{\frac{1}{2}}(\sigma_{33})^{\frac{1}{2}}} \exp\left[-\frac{1}{2\sigma_{33}} e_{3i}^2\right] \\ \cdot F\left[\frac{x_{1i}\beta_1 + (\sigma_{13}/\sigma_{33})e_{3i}}{((\sigma_{33} - \sigma_{23}^2)/\sigma_{33})^{1/2}}, \frac{x_{2i}\beta_2 + (\sigma_{23}/\sigma_{33})e_{3i}}{((\sigma_{33} - \sigma_{23})/\sigma_{33})^{1/2}}, \frac{\rho\sigma_{33} - \sigma_{13}\sigma_{23}}{(\sigma_{33} - \sigma_{13})^{\frac{1}{2}}(\sigma_{33} - \sigma_{23})^{\frac{1}{2}}}\right] \\ /F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)$$

(See APPENDIX D.) Then, the conditional distribution of y_{3i} given its observability is:

$$\begin{aligned}
 (13) \quad & h(y_{3i} | y_{1i}^* \geq 0, y_{2i}^* \geq 0) \\
 &= \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma_{33})^{\frac{1}{2}}} \exp \left[-\frac{1}{2\sigma_{33}} (y_{3i} - x_{3i}\beta_3)^2 \right] \\
 &\cdot F \left[\frac{x_{1i}\beta_1 + (\sigma_{13}/\sigma_{33})(y_{3i} - x_{3i}\beta_3)}{((\sigma_{33} - \sigma_{23}^2)/\sigma_{33})^{1/2}}, \frac{x_{2i}\beta_2 + (\sigma_{23}/\sigma_{33})(y_{3i} - x_{3i}\beta_3)}{((\sigma_{33} - \sigma_{23}^2)/\sigma_{33})^{1/2}}, \right. \\
 &\quad \left. \frac{\rho\sigma_{33} - \sigma_{13}\sigma_{23}}{(\sigma_{33} - \sigma_{13}^2)^{1/2} (\sigma_{33} - \sigma_{23}^2)^{1/2}} \right] / F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)
 \end{aligned}$$

The log-likelihood function is:

$$\begin{aligned}
 (14) \quad & \ln L^{II}(\beta_1, \beta_2, \beta_3, \rho, \sigma_{13}, \sigma_{23}, \sigma_{33}) \\
 &= \sum_{i=1}^N \{ y_{1i}y_{2i} \ln[h(y_{3i} | y_{1i}^* \geq 0, y_{2i}^* \geq 0) \Pr(y_{1i}^* \geq 0, y_{2i}^* \geq 0)] \\
 &\quad + y_{1i}(1-y_{2i}) \ln[\Pr(y_{1i}^* \geq 0, y_{2i}^* < 0)] \\
 &\quad + (1-y_{1i}) \ln[\Pr(y_{1i}^* < 0)] \} \\
 &= \sum_{i=1}^N \{ y_{1i}y_{2i} [-(1/2) \ln 2 - (1/2) \ln \sigma_{33} - (1/2\sigma_{33})(y_{3i} - x_{3i}\beta_3)^2 \\
 &\quad + \ln(H_i)] + y_{1i}(1-y_{2i}) \ln(\Phi_{1i} - F_i) \\
 &\quad + (1-y_{1i}) \ln(1 - \Phi_{1i}) \}
 \end{aligned}$$

where,

$$\begin{aligned}
 H = F \left[\frac{x_{1i}\beta_1 + (\sigma_{13}/\sigma_{33})(y_{3i} - x_{3i}\beta_3)}{((\sigma_{33} - \sigma_{23}^2)/\sigma_{33})^{1/2}}, \frac{x_{2i}\beta_2 + (\sigma_{23}/\sigma_{33})(y_{3i} - x_{3i}\beta_3)}{((\sigma_{33} - \sigma_{23}^2)/\sigma_{33})^{1/2}}, \right. \\
 \left. \frac{\rho\sigma_{33} - \sigma_{13}\sigma_{23}}{(\sigma_{33} - \sigma_{13}^2)^{1/2} (\sigma_{33} - \sigma_{23}^2)^{1/2}} \right]
 \end{aligned}$$

$$F_i = F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)$$

$$\Phi_{1i} = \Phi(x_{1i}\beta_1).$$

If $\sigma_{13} = \sigma_{23} = 0$, the log-likelihood function given in (14)

becomes:

$$\begin{aligned}
 (15) \quad \ln L(\beta_1, \beta_2, \beta_3, \rho, \sigma_{33}) \\
 = -(N_1/2) \ln 2 - (N_1/2) \ln \sigma_{33} - (1/2\sigma_{33}) \sum_{i=1}^{N_1} (Y_{3i} - x_{3i}\beta_3)^2 \\
 + \ln L^C(\beta_1, \beta_2, \rho)
 \end{aligned}$$

where $\ln L^C$ is given in (6). (15) implies that $(\beta_3', \sigma_{33})'$ and $(\beta_1', \beta_2', \rho)'$ can be estimated separately. $(\beta_3', \sigma_{33})'$ can be efficiently estimated by applying the OLS to (4.3), and maximizing (6) provides the ML estimator of $(\beta_1', \beta_2', \rho)'$. As we see in (12), if $\sigma_{13} = \sigma_{23} = 0$,

$$\begin{aligned}
 (16) \quad g(e_{3i} | e_{1i} \geq -x_{1i}\beta_1, e_{2i} \geq -x_{2i}\beta_2) \\
 = \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma_{33})^{\frac{1}{2}}} \exp \left[-\frac{1}{2\sigma_{33}} e_{3i}^2 \right]
 \end{aligned}$$

This shows that the conditional distribution of e_{3i} is normal, and that all e_{3i} 's for $i=1, 2, \dots, N_1$ are homoscedastic. That is why the OLS estimator of $(\beta_3', \sigma_{33})'$ is efficient if $\sigma_{13} = \sigma_{23} = 0$.

One thing is noteworthy. In Model II (the Reduced-Form model) there is no sample selection bias if $\sigma_{13} = \sigma_{23} = 0$. In Model I (the Structural model) this means that $\Sigma_{33} - \Sigma_{13} = \Sigma_{23} = 0$. If $\Sigma_{33} = 0$, the wage equation given in (1.3) becomes deterministic, which is difficult to believe. Also, it may not be usual that $\Sigma_{33} = \Sigma_{13}$. Therefore, Model I almost surely has a selectivity problem.

All the structural parameters in Model I also can be directly estimated by the ML estimation method. We can easily derive the log-likelihood function for Model I:

$$\begin{aligned}
(17) \quad \ln L^I(\delta_1, \delta_2, \delta_3, \Sigma) \\
= \sum_{i=1}^{N_1} \{ Y_{1i} Y_{2i} [-(1/2) \ln 2 - (1/2) \ln \sigma_{33} \\
- (1/2 \sigma_{33}) (w_i - z_{3i} \delta_3)^2 + H_i] + Y_{1i} (1 - Y_{2i}) \ln(\Phi_{1i} - F_i) \\
+ (1 - Y_{1i}) \ln(1 - \Phi_{1i}) \}
\end{aligned}$$

where

$$\rho = (\Sigma_{23} - \Sigma_{12}) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2}$$

$$\sigma_{13} = (\Sigma_{33} - \Sigma_{13}) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2}$$

$$F_i = F(x_{1i}\beta_1, z_{3i}\delta_3, \rho)$$

$$\begin{aligned}
H = F \left[\frac{x_{1i}\beta_1 + (\sigma_{13}/\Sigma_{33})(w_i - z_{3i}\delta_3)}{((\Sigma_{33} - \sigma_{13}^2)/\Sigma_{33})^{1/2}}, \quad \frac{z_{2i}\delta_2 + (\Sigma_{23}/\Sigma_{33})(w_i - z_{3i}\delta_3)}{((\Sigma_{33} - \Sigma_{23}^2)/\Sigma_{33})^{1/2}}, \right. \\
\left. \frac{\rho \Sigma_{33} - \sigma_{13} \Sigma_{23}}{(\Sigma_{33} - \sigma_{13}^2)^{1/2} (\Sigma_{33} - \Sigma_{23}^2)^{1/2}} \right]
\end{aligned}$$

$$F_i = F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)$$

$$\Phi_{1i} = \Phi(x_{1i}\beta_1)$$

$$x_{1i}\beta_1 = (z_{3i}\delta_3 - z_{1i}\delta_1) / (\Sigma_{11} + \Sigma_{33} - 2\Sigma_{13})^{1/2}$$

Maximizing (17) with respect to $\delta_1, \delta_2, \delta_3$, and Σ generates the ML estimates of the parameters. For identification at least one variable in z_{3i} must not be included in z_{1i} .

III. Data

The sample of married women is taken from the University of Michigan's Panel Study of Income Dynamics (PSID) for 1981. These women are between the ages of 18 and 60 years. The sample excludes wives who are in the agricultural sector; self-employed; retired; disabled; students; or not in the continental U.S.. Wives whose total family money income in

1980 is less than \$5,000 are also excluded from the sample. Black, White, and Hispanics are included in the sample, but people of other races are eliminated. Some respondents' answers to various question items are inconsistent. For example, some wives are reported as working in 1981, but are recorded as having zero hourly wage rates. Those observations with unreliable answers are also excluded. After this process, the sample contains 1962 observations. Of these, 923 people are recorded as working at the time the survey was taken in 1981, and are therefore categorized as the employed. 956 women are reported as housewives, and they belong to the NLF group. This means that 48.7% of women in the data are out of the labor force. 83 women, or 8.2% of those in the labor force, are looking for jobs or temporarily laid-off. These women are regarded as the unemployed.

The definitions, means, and standard deviations of the variables used in the analysis are shown in Table 1. The mean hourly wage rate shown in Table 1 is quite low. This is misleading because more than half of the women in the sample are not currently employed and their "wage" is recorded zero. If we consider only the employed, the average hourly wage is \$5.87. Other earned and unearned income (OFINC) could affect a wife's labor-force participation and ability to get acceptable job offers. This variable is obtained by extracting the wife's labor income from total family money income. Since the logs of reservation wage and market wage are used in the model, I choose the log of OFINC (LOFINC) as

a regressor. Some families in the data have zero OFINC. Therefore, one is added to OFINC to calculate LOFINC; i.e., $LOFINC = \ln(OFINC + 1)$.

Regional effects are captured by city size and area of residence. The dummy variables, URB and REGS represent residency in an SMSA and in the South, respectively. Regional effects could be investigated in more detail if more dummy variables for regions were created. However, in this case, a proportionate increase in computational cost would follow. Demographic variables, such as years of education (ED), age (AGE), the number of children below the age of 6 years (KIDS), and a dummy variable for race (MINOR), are also used. Blacks and Hispanics are grouped as a single minority. Work experience could affect the market wage an individual can earn once employed or her job-match skill. The actual number of years worked since the age of 18 (EXP) is used to capture this effect. Finally, this study includes the local unemployment rate in order to capture differing demand conditions across areas. The PSID reports the unemployment rate in the respondent's county. This variable (UNEMPR) is used as the local unemployment rate.

The explanatory variables in the equation for labor-force decisions are the constant; ED; URB; MINOR; REGS; UNEMPR; LOFINC; KIDS; AGE, and AGE squared divided by 1,000 (AGE2); EXP, and EXP squared divided by 1,000 (EXP2). The vector of explanatory variables in the job-match equation includes the same variables in the labor-force decision except AGE and

AGE2. The explanatory variables in the wage equation are all the same variables as in the labor-force decision except LOFINC and KIDS.

IV. Empirical Results.

The first column of Table 2 reports the estimates of a simple probit model for the No-Friction model. The last two columns describe the results of the simple probit models for the Friction model, assuming zero correlation between error terms in the labor-force and employment decisions. Both the first and second columns are about willingness to work. However, the results in these two columns are derived by different treatment of the unemployed. The model for no friction does not regard those unemployed people as having a desire to work, while the model for friction does. In spite of these differences, the estimates in the second column are generally similar to those in the first column and the sizes of effects and their signs satisfy our expectation. This similarity may come from the fact that only 4.2% of women in the data are unemployed, so that their treatment will not change the results dramatically.

Some differences exist, though, between the results in the first and the second columns. First, the effect of race is four times as large in the second column compared to that in the first column. In fact, MINOR is insignificant when the unemployed are considered as preferring not to work. On the contrary, as we see in the second column, when we interpret

the unemployed as willing to work, MINOR becomes significant at the 10% level. This means that the model for no friction understates the effect of race on the preference to work. Second, the coefficient of UNEMPR is -3.77 (significant at the 1% level) under the assumption of no friction while it is -2.38 (significant at the 10% level) under the assumption of frictions in labor market. Therefore, the No-Friction model seems to exaggerate the effect of the local unemployment rate on the willingness to work.

The third column in Table 2 shows the estimates of parameters for the employment status equation based on the Friction model. MINOR has a significantly negative effect on employment status. This confirms our expectation. Blacks and other minorities are more willing to enter the labor force, but either their job-search ability is less than that of whites, or there is discrimination in employment. This may also explain why the No-Friction model underestimates the effect of race on the preference for work. Under the Friction model, MINOR has two opposite effects on employment status. Minorities have higher probabilities of being employed because they are more likely to be in the labor force. At the same time, they are more likely to be unemployed because of their poor job-search skills. These two opposite effects are captured by one equation under the No-Friction model, and therefore, these opposite effects cancel out. The Friction model can capture those different effects of race separately by two different equations.

The other notable result in the second column is that the local unemployment rate has a huge effect on employment status. This implies that the demand side of the labor market is a major factor determining an individual's employment status. The effects on employment status of the local unemployment rate are twofold. First, as we see in the third column, a higher unemployment rate decreases the probability of an individual being in labor force. Second, once she enters the labor market, a woman has a lower probability of being employed. These two different negative effects of the unemployment rate are captured by the single equation for preference for work under the model for no friction. This may explain why the Friction model generates the exaggerated effect of UNEMPR on preference to work.

Even though the Friction Model explains an individual's behavior more completely than the No-Friction Model, there is no direct method for discriminating between the two models, because one is not nested in the other. One roundabout way is to compare the goodness-of-fit of the different models. In probabilistic-choice models, the proportion of successful predictions of the choices made is widely used for the measure of goodness-of-fit. (See Maddala [1987], pp 76 - 77.) Table 3 describes the frequencies of predicted outcomes for the No-Friction Model, while Tables 4-A and -B show those for the Friction Model. In Table 3, 69.7% of predicted outcomes are correct. According to Tables 4-A and -B, the Friction Model correctly predicts 76.1% of the total outcomes. The Friction

Model shows better predictive power than the No-Friction Model.

In short, as we see in Table 2, the Friction model explains an individual's labor-force status in a more complete way than the No-Friction model. Some variables have opposite effects on employment status, leading the No-Friction model to understate the effect of those variables on the willingness to work. On the contrary, if some variables affect employment directly and indirectly through the labor-force decision but in the same direction, the No-Friction model overestimates their effects on the willingness to work. These results imply that recognizing friction in the labor market can provide a more reliable explanatory mechanism for the supply approach to analyzing employment status.

Until now, we have assumed zero correlation coefficient between labor-force and employment decisions. Table 5 reports results for the censored probit model, allowing a non-zero correlation coefficient. A test of zero correlation yields the LR statistic of 4.89, larger than the critical $\chi^2(1)$ value of 3.84 at the 5% level. Also, the conventional t-test shows significance of ρ at the 1% level. In spite of this fact, the results in Table 5 are generally close to those in the last two columns of Table 2. This is not surprising, because the estimates in Table 2 obtained under the assumption of zero correlation are still consistent. Some small differences also follow. If we allow ρ to be different from zero, the signs of some of the coefficients in the employment decision equation

become more reasonable. For example, compared with those in the third column of Table 2, the estimated coefficients of ED and LOFINC become positive and negative, respectively, following our expectation. MINOR has an insignificant coefficient in Table 5, but the sign of the coefficient is still negative. The number of years of work experience has no significant explanatory power for employment status, when zero correlation is assumed. However, allowing non-zero correlation reveals significant effects of work experience and the expected inverted-U shape. Finally, we can test the hypothesis that the coefficients entering in both labor-force and employment decisions have the same sizes of effects on those decisions. Under this null hypothesis, the restricted log-likelihood function (results not presented) has the value of -1421.76. Then the value of the LR statistic is 60.50, which is considerably larger than the critical $\chi^2(10)$ value of 23.21 at the 1% level. This implies that the explanatory variables affect labor-force and employment status separately in significantly different ways.

Jointly estimating labor-force and employment decisions and the wage equation could generate more efficient estimates of the parameters describing labor-force and employment decisions. This could be done based on the Reduced-Form model given by (4.1)-(4.3). Table 6 shows the ML estimation results for this joint estimation. Compared with the results in Table 5 derived by the censored probit model, many estimates of the parameters describing employment status become significant, or

more significant. For example, ED, URB, MINOR, and LOFINC have no significant effects on employment status when they are estimated by the censored probit model. (See Table 5) They become significant at the 5% level when the joint model is estimated, as we see in Table 6. This is due to the increased efficiency of MLE for the joint model as compared to the censored probit model. The estimates in the first and the second columns of Table 6 are generally close to their counterparts in Table 5. MINOR has an insignificant effect on labor-force status in Table 6 while having a significant effect in Table 5. However, the sizes of the effect of MINOR described in the two tables are not substantially different.

Almost all the estimates in Table 6 have the signs we would expect. More educated people and residents of bigger cities have a higher probability of being in the labor force, and are more likely to be employed. They also obtain higher wage rates once employed. Minorities are more willing to participate, but their probability of being unemployed is higher. Even when they find jobs, their wage rates are lower than those of whites. Age has no significant effect on labor-force status or the market wage rate. (This is because experience is included in the equations.) The local unemployment rate seems not to affect labor-force status significantly. However, the local unemployment rate has a large significant effect on employment status. A higher unemployment rate generates a greater likelihood of an individual being unemployed, and substantially decreases the

hourly wage rate. The income earned by other family members decreases a married woman's willingness to work and her probability of being employed. This negative effect of other family members' income on a woman's employment status is not surprising in light of economic search models and commonsense.

A woman with higher income due to others' earning will be more selective in choosing jobs.

The number of children has very significant negative effects on both labor-force and employment decisions, as we would expect. This effect also could be explained by a search framework. A woman with more children will have more burdens of housework, which will lessen her intensity of job search. Years of work experience have quadratic effects on LF and employment status and the market wage rate. According to the estimates in Table 6, more experienced women have a greater desire to work, a higher likelihood of being employed, and a higher wage rate.

Table 6 also shows that the error terms for labor-force and employment decisions are significantly correlated with the error term for the market wage rate. The covariances between errors in the labor-force decision and the wage rate, and between errors in employment status and the wage rate, are positive and significant at the 1% level. The same unobserved individual characteristics that make a woman more likely to enter the labor force and to find a job also tend to lead to an expectedly higher wage.

This result also implies the existence of sample selection bias in the estimation of an equation for observed wage rates. Table 7 shows the results of the restricted MLE under the null hypothesis of no selection bias. The restricted MLE of the parameters in the wage equation are just OLS applied to the observations on the employed women, while the restricted MLE for labor-force and employment decisions are identical to the estimates obtained by the censored probit model. Therefore, the results of Table 5 are repeated in the first and the second columns of Table 7. A test of no selection bias yields an LR statistic of 24.40, considerably larger than the critical value for $\chi^2(2)$ of 9.21 at the 1% level. This suggests that the OLS estimates could be seriously biased. Comparing the third column of Table 6 with that of Table 7, we can see that OLS underestimates the effect of the local unemployment rate on the wage rate. Under the null hypothesis of no selection bias, the estimated coefficient of UNEMPR is insignificant, while the unrestricted MLE shows a significant and large negative effect of UNEMPR on the wage rate. The OLS estimates also understate the effect of work experience. The absolute values of the coefficients of EXP and EXP2 from the unrestricted MLE are almost twice as big as those obtained by OLS.

Sample selection biases can be corrected by the two-stage estimation method. Table 8 reports the results. The estimated covariance between the labor-force decision and the wage rate is significant and has a positive sign, while the

covariance between employment status and the wage rate is insignificant and has a negative sign. The F-statistic, which is asymptotically equivalent to the LM statistic, has a value of 8.46, greater than the critical F value of 4.61. Again, the null hypothesis of no selection bias is rejected. The two-stage estimation method corrects the biases of the OLS estimates for EXP and EXP2 quite well, and the estimated effects of EXP are quite close to those obtained using the unrestricted MLE. However, some estimates have an unexpected sign. Even though AGE and AGE2 are not significant in either the two-stage or the MLE estimates, in the two-stage estimates older women are predicted to receive lower wage rates. This is an unbelievable result. The estimated coefficient of UNEMPR is also insignificant and has an unexpected positive sign. Therefore, the two-stage estimates seem to fail in successfully eliminating biases in the OLS estimates. Also, the two-stage method produces an insignificant and negative covariance between errors in employment status and the wage rate.

Compared to the unrestricted MLE results on the effect of race, the two-stage method underestimates the effect on the wage rate. The estimated coefficient of MINOR is not significantly different from zero. Considering many studies showing wage discrimination, we expect a significant and negative effect of minority on wage rates, while we get by MLE but not by the two stage method. On this point, the MLE result seems to be more reliable.

In short, the two-stage estimation method, as well as the MLE method, confirm the presence of sample selection biases. However, compared with the MLE, the two-stage method does not successfully eliminate the biases of the OLS estimates, and in some cases generates perversely signed coefficients. These results imply the superiority of MLE over the two-stage method.

As a final step, Table 9 describes the result of the joint estimation of the employment decision and the reservation and market wage rates, which is obtained by applying unrestricted MLE to the Structural model given in (1.1)-(1.3). UNEMPR, EXP, and EXP2 are excluded from the reservation wage equation. This is done for a computational reason. For the identification of the equation for the reservation wage, at least one explanatory variable in the wage equation must be excluded from the set of explanatory variables of the reservation-wage equation. The reservation wage could be interpreted as a woman's value of leisure or the nonmarket value of housework. In this sense, there is no reason why UNEMPR, EXP, and EXP2 should affect the reservation wage. Also, the exclusion of these variables can be justified by their insignificance in explanation of reservation wages. If we include the variables, MLE on the Structural model amounts to that on the Reduced-Form model. The log-likelihood value given in Table 9 is not significantly different from that in Table 6, rejecting the hypothesis of significant effects of UNEMPR, EXP, and EXP2 on the reservation wage.

The estimated coefficients of the explanatory variables in the employment decision and the market wage rate equation are almost identical to their counterparts in Table 6. Error terms in the reservation wage rate equation are significantly and positively correlated with those in the market wage rate equation (see Σ_{13} in Table 9), but are not significantly correlated with those in the employment decision equation. A woman with an unexpectedly high reservation wage rate tends to get an unexpectedly high wage rate when she is employed, while the unexplained part of her reservation wage rate does not affect the probability of her being employed.

The results for the reservation wage equation are shown in the first column of Table 9. More educated people have a higher reservation wage rate. This, however, does not mean that more educated people are less willing to work, as an extra year of education increases the market wage more than the reservation wage. The high labor-force participation rate of minorities can be also explained by the results in the first column showing their lower reservation wage. The effect of age on reservation wage is insignificant but positive. As we expected, the more other income in a family, and the more children the family has, the higher the wife's reservation wage rate. The elasticity of the reservation wage rate with respect to other income is 0.15. Table 9 also shows that residents in SMSAs or the South have higher reservation wages. There is no theory which explains the effects of region on married women's reservation wages. However, the LR test of no

regional effects yields a χ^2 statistic of 117.95, considerably above the $\chi^2(2)$ critical value of 9.21 at the 1% level. This is a quite interesting result, and further examination seems to be required.

V. Conclusion

This paper presents a joint estimation method for labor-force and employment decisions and market wage. This is based on a Full Information Maximum Likelihood procedure as well as on a two-step method. Frictions in the labor market are assumed, and therefore the unemployed are recognized as behaviorally different from non-participants. The information about unemployed workers present in a given sample can be used to estimate the parameters describing the probability of a particular individual's being employed. This joint estimation method provides an explanatory mechanism showing the different and separate effects a variable could have on labor-force participation and employment decisions, as well as improving the estimates of parameters in the wage equation.

The traditional labor supply model, which assumes no friction in the labor market, does not explain married women's employment status in a satisfactory way. Some variables could affect employment decisions directly and indirectly through preference for work. The No-Friction model, assuming that a person's employment status depends only on that person's willingness to work, cannot discriminate between those two different effects, so that it usually generates biased

estimates of parameters describing preferences for work.

Compared to other methods for estimating labor-force and employment decisions jointly (for example, bivariate probit methods), the estimation procedure including the wage equation generates more significant and reasonably signed estimates. A given data set usually contains a relatively small number of unemployed people, and therefore the information on them may not be enough to generate much more efficient estimates of parameters in employment decisions. Jointly considering the wage rate will be helpful for more efficient estimation of parameters in labor-force and employment decisions. These two decision rules can censor the observed distribution of wage rates, if error terms in the two equations describing labor-force and employment decisions are correlated with those in the wage equation. Therefore, as this paper shows, the information on wage rates can improve the efficiency of estimates of parameters explaining labor-force and employment decisions.

Table 1. Means, Standard Deviations and Definitions of Variables

| Variables | Definition | Mean | S.D. |
|------------------|----------------------------------------------------|-------------|-------------|
| EMP | Employed=1 | 0.4704 | 0.4993 |
| NLF | NLF=1 | 0.4873 | 0.5000 |
| WRATE | Hourly wage rate(\$) | 2.7615 | 3.3966 |
| LRATE | Log of WRATE | 0.8002 | 0.8758 |
| ED | Years of Education | 12.205 | 2.2735 |
| URB | Resident in SMSA=1 | 0.6865 | 0.4640 |
| MINOR | Nonwhite=1 | 0.2717 | 0.4449 |
| AGE | Years of Age | 37.485 | 11.298 |
| AGE2 | AGE ² /1000 | 1.5327 | 0.8973 |
| REGS | South=1 | 0.3435 | 0.4750 |
| UNEMPR | Unemployment rate in the resident's county in 1980 | 0.0743 | 0.0242 |
| OFINC | Other family member's income in 1980(\$) | 24850.7 | 20372.2 |
| LOFINC | Log of (OFINC+1) | 9.9052 | 0.6822 |
| KIDS | Number of children ≤ 5 years of age | 0.5076 | 0.7745 |
| EXP | Number of years worked since age 18 | 8.3578 | 6.9111 |
| EXP2 | EXP ² /1000 | 0.1176 | 0.1912 |

Table 2. Simple Probit Models for Labor-Force Participation and Employment

| Model | No Friction | Friction | |
|-----------|------------------------|------------------------|------------------------|
| Dep. Var. | EMP | LF | EMP |
| CONSTANT | 1.9064*** (0.6564) | 2.8483*** (0.6720) | 0.7660 (0.8787) |
| ED | 0.0688*** (0.0151) | 0.0797*** (0.0156) | -0.0099 (0.0299) |
| URB | 0.1793*** (0.0693) | 0.1403** (0.0699) | 0.2031 (0.1369) |
| MINOR | 0.0403 (0.0776) | 0.1615** (0.0781) | -0.3846** (0.1700) |
| AGE | 0.0022 (0.0262) | -0.0252 (0.0265) | - |
| AGE2 | -0.4820 (0.3187) | -0.2201 (0.3219) | - |
| REGS | 0.1117 (0.0729) | 0.1008 (0.0730) | 0.1750 (0.1827) |
| UNEMPR | -3.7654*** (1.3442) | -2.3757* (1.3645) | -7.7640*** (2.6920) |
| LOFINC | -0.2713*** (0.0575) | -0.3224*** (0.0603) | 0.1147 (0.0885) |
| KIDS | -0.4954*** (0.0383) | -0.5131*** (0.0386) | -0.2054** (0.0941) |
| EXP | 0.1371*** (0.0176) | 0.1543*** (0.0177) | 0.0332 (0.0334) |
| EXP2 | -2.3910*** (0.5982) | -2.8442*** (0.6012) | -0.4396 (1.4115) |
| N | 1962 | 1962 | 1006 |
| Log L | -1148.8 | -1128.9 | -265.01 |

* significant at the 10% level
 ** significant at the 5% level
 *** significant at the 1% level
 standard errors in parenthesis

Table 3. Frequencies of Actual and Predicted Outcomes for the Probit Model of No-Friction.

| Actual\Predicted | EMP=0 | EMP=1 | Total |
|-------------------------|--------------|--------------|--------------|
| EMP = 0 | 761 | 278 | 1039 |
| EMP = 1 | 316 | 607 | 923 |
| Total | 1077 | 885 | 1962 |

Table 4-A. Frequencies of Actual and Predicted Outcomes for the Probit Model of Labor-Force Status Based on the assumption of Frictions in Labor Market

| Actual\Predicted | LF=0 | LF=1 | Total |
|------------------|------|------|-------|
| LF = 0 | 636 | 320 | 936 |
| LF = 1 | 276 | 730 | 1006 |
| Total | 912 | 1050 | 1962 |

Table 4-B. Frequencies of Actual and Predicted Outcomes for the Probit Model of EMP Status Based on the Assumption of Frictions in Labor Market

| Actual\Predicted | EMP=0 | EMP=1 | Total |
|------------------|-------|-------|-------|
| EMP = 0 | 0 | 83 | 83 |
| EMP = 1 | 0 | 923 | 923 |
| Total | 0 | 1006 | 1006 |

Table 5. Censored Probit Estimates of the Friction Model

| Dep. Var. | LF | EMP |
|-----------|------------------------|------------------------|
| CONSTANT | 2.9638*** (0.6659) | 1.0227*** (0.8859) |
| ED | 0.0793*** (0.0157) | 0.0363*** (0.0321) |
| URB | 0.1437** (0.0703) | 0.2240** (0.1232) |
| MINOR | 0.1645** (0.0788) | -0.2469 (0.1600) |
| AGE | -0.0356 (0.0259) | - |
| AGE2 | -0.0941 (0.3134) | - |
| REGS | 0.1014 (0.0735) | 0.1876 (0.1585) |
| UNEMPR | -2.2564* (1.3726) | -7.7352*** (2.3775) |
| LOFINC | -0.3161*** (0.0602) | -0.0465 (0.1180) |
| KIDS | -0.5164*** (0.0393) | -0.3668*** (0.0947) |
| EXP | 0.1566*** (0.0177) | 0.0640** (0.0301) |
| EXP2 | -2.9305*** (0.6015) | -1.1999*** (1.2245) |
| ρ | | 0.6159*** (0.2100) |
| N | 1962 | 1006 |
| Log L | | -1391.51 |

* significant at the 10% level
 ** significant at the 5% level
 *** significant at the 1% level
 standard errors in parenthesis

Table 6. Joint Estimation of Labor-Force, Employment Decisions and the Wage Equation (the Reduced-Form model)

| Dep. Var. | LF | EMP | LWAGE |
|-----------------------------------------|------------------------|------------------------|------------------------|
| CONSTANT | 3.5160*** (0.6600) | 2.2994*** (0.8264) | 0.1640 (0.1949) |
| ED | 0.0884*** (0.0158) | 0.0698** (0.0285) | 0.0764*** (0.0490) |
| URB | 0.1547** (0.0699) | 0.2559** (0.1163) | 0.1910*** (0.0289) |
| MINOR | 0.1057 (0.0802) | -0.3364** (0.1592) | -0.0959*** (0.0289) |
| AGE | -0.0213 (0.0258) | - | 0.0079 (0.0104) |
| AGE2 | -0.2192 (0.3132) | - | -0.1931 (0.1250) |
| REGS | 0.0746 (0.0742) | 0.0967 (0.1476) | -0.0629** (0.0305) |
| UNEMPR | -2.1860 (1.3617) | -7.1815*** (2.2748) | -1.1938** (0.4973) |
| LOFINC | -0.4172*** (0.0597) | -0.2473** (0.1066) | - |
| KIDS | -0.4565*** (0.0391) | -0.2583*** (0.0877) | - |
| EXP | 0.1530*** (0.0180) | 0.0783*** (0.0279) | 0.0493*** (0.0080) |
| EXP2 | -2.8690*** (0.6213) | -1.2782 (1.1404) | -0.8524*** (0.2055) |
| ρ | 0.6858*** (0.1833) | | |
| $\sigma_{13}, \sigma_{23}, \sigma_{33}$ | 0.2215*** (0.0515) | 0.2922*** (0.0353) | 0.1442*** (0.0127) |
| N | 1962 | 1006 | 923 |
| Log L | | -1664.86 | |

** significant at the 5% level
 *** significant at the 1% level
 standard errors in parenthesis

Table 7. Restricted MLE Estimates of the Three-Equation System ($\sigma_{13}=\sigma_{23}=0$)

| Dep. Var. | LF | EMP | LWAGE |
|-----------|------------------------|------------------------|------------------------|
| CONSTANT | 2.9638*** (0.6659) | 1.0227*** (0.8859) | 0.3834** (0.1899) |
| ED | 0.0793*** (0.0157) | 0.0363*** (0.0321) | 0.0694*** (0.0053) |
| URB | 0.1437** (0.0703) | 0.2240** (0.1232) | 0.1696*** (0.0253) |
| MINOR | 0.1645** (0.0788) | -0.2469 (0.1600) | -0.0928*** (0.0276) |
| AGE | -0.0356 (0.0259) | - | 0.0134 (0.0097) |
| AGE2 | -0.0941 (0.3134) | - | -0.2147* (0.1203) |
| REGS | 0.1014 (0.0735) | 0.1876 (0.1585) | -0.0809*** (0.0260) |
| UNEMPR | -2.2564* (1.3726) | -7.7352*** (2.3775) | -0.5296 (0.4771) |
| LOFINC | -0.3161*** (0.0602) | -0.0465 (0.1180) | - |
| KIDS | -0.5164*** (0.0393) | -0.3668*** (0.0947) | - |
| EXP | 0.1566*** (0.0177) | 0.0640** (0.0301) | 0.0273*** (0.0062) |
| EXP2 | -2.9305*** (0.6015) | -1.1999*** (1.2245) | -0.4301* (0.1888) |
| ρ | 0.6159*** (0.2100) | | |
| N | 1962 | | |
| Log L | -1391.51 | | |
| | 923 | | |
| | -285.55 | | |

* significant at the 10% level
 ** significant at the 5% level
 *** significant at the 1% level
 standard errors in parenthesis

Table 8. Results for Two-Stage Estimation of the Wage Equation

| Dependent variable | LWAGE |
|--------------------|------------------------|
| Variables | |
| CONSTANT | 0.4133*** (0.2475) |
| ED | 0.0769*** (0.0070) |
| URB | 0.1504*** (0.0375) |
| MINOR | -0.0178 (0.0516) |
| AGE | -0.0020 (0.0119) |
| AGE2 | -0.1283 (0.1382) |
| REGS | -0.0894** (0.0369) |
| UNEMPR | 0.2323 (0.8870) |
| EXP | 0.0505*** (0.0106) |
| EXP2 | -0.8612*** (0.2792) |
| σ_{13} | 0.2553*** (0.0949) |
| σ_{23} | -0.3613 (0.2556) |
| σ_{33} | 0.2042 |
| R ² | 0.3321 |
| N | 923 |

* significant at the 10% level

** significant at the 5% level

*** significant at the 1% level

correct standard errors in parenthesis.

Table 9. Unrestricted Joint Estimation of Employment Decisions and Reservation and Market Wage Rates (the Structural Model)

| Dependent Variable | Log of Unobserved Reservation Wage Rate | EMP | LWAGE |
|-----------------------------------------|-----------------------------------------|------------------------|------------------------|
| Variable | | | |
| CONSTANT | -1.5060*** (0.2770) | 2.2593*** (0.8332) | 0.1743 (0.1930) |
| ED | 0.0451*** (0.0075) | 0.0679** (0.0289) | 0.0764*** (0.0049) |
| URB | 0.1376*** (0.0326) | 0.2594** (0.1164) | 0.1918*** (0.0289) |
| MINOR | -0.1336*** (0.0382) | -0.3410** (0.1601) | -0.0965*** (0.0314) |
| AGE | 0.0116 (0.0112) | --- | 0.0056 (0.0101) |
| AGE2 | -0.0710 (0.1416) | --- | -0.1668 (0.1217) |
| REGS | -0.0853** (0.1485) | 0.1032 | -0.0610** (0.0303) |
| UNEMPR | --- | -7.0411*** (2.2617) | -0.9428** (0.3994) |
| LOFINC | 0.1473*** (0.0284) | -0.2409** (0.1080) | --- |
| KIDS | 0.1599*** (0.0220) | -0.2568*** (0.0882) | --- |
| EXP | --- | 0.0786*** (0.0278) | 0.0518*** (0.0075) |
| EXP2 | --- | -1.3000 (1.1410) | -0.9316*** (0.1838) |
| $\Sigma_{11}, \Sigma_{12}, \Sigma_{13}$ | 0.1093*** (0.0210) | 0.0510 (0.0745) | 0.0652*** (0.0191) |
| Σ_{23}, Σ_{33} | | 0.2890*** (0.0366) | 0.1446*** (0.0127) |
| N | | 1962 | |
| Log L | | -1665.33 | |

* significant at the 10% level
 ** significant at the 5% level
 *** significant at the 1% level
 standard errors in parenthesis

APPENDICES

APPENDIX A

In Model II,

$$(A.1) \quad E(e_3 | e_1 \geq -x_1\beta_1, e_2 \geq -x_2\beta_2) = \sigma_{13}\mu_1 + \sigma_{23}\mu_2$$

$$(A.2) \quad E(e_3^2 | e_1 \geq -x_1\beta_1, e_2 \geq -x_2\beta_2) = \sigma_{33} - \sigma_{13}^2\mu_1 - \sigma_{23}^2\mu_2 \\ - (\rho\sigma_{13}^2 - 2\sigma_{13}\sigma_{23} + \rho\sigma_{23}^2)\mu_3$$

(For simplicity, the subscript i is suppressed.)

Proof. Let $f(e_1, e_2, e_3)$ be the trivariate normal distribution function of $(e_1, e_2, e_3)'$ where $(e_1, e_2, e_3)'$ has zero mean and covariance matrix Ω given below; and let $f(e_1, e_2, \rho)$ be the standard bivariate normal distribution function with correlation coefficient, ρ . Let

$$\Omega = \begin{pmatrix} 1 & \rho & \sigma_{13} \\ . & 1 & \sigma_{23} \\ . & . & \sigma_{33} \end{pmatrix}; \quad \Omega^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} & \Omega^{13} \\ & \Omega^{22} & \Omega^{23} \\ & & \Omega^{33} \end{pmatrix}$$

We can easily show that

$$(A.3) \quad E(e_3 | e_1 \geq -x_1\beta_1, e_2 \geq -x_2\beta_2) = \frac{1}{F(x_1\beta_1, x_2\beta_2, \rho)} \\ \cdot \int_{-x_1\beta}^{\infty} \int_{-x_2\beta}^{\infty} \int_{-\infty}^{\infty} e_3 f(e_1, e_2, e_3) de_3 de_2 de_1$$

Note that

$$f(e_1, e_2, e_3) = \frac{1}{(2\pi)^{3/2} |\Omega|^{1/2}} \exp\left\{-\frac{1}{2} (\Omega^{11}e_1^2 + \Omega^{22}e_2^2 \right. \\ \left. + \Omega^{33}e_3^2 + 2\Omega^{12}e_1e_2 + 2\Omega^{13}e_1e_3 + 2\Omega^{23}e_2e_3)\right\} \\ = \frac{1}{(2\pi)^{3/2} |\Omega|^{1/2}} \exp\left\{-\frac{\Omega^{33}}{2} \left(e_3 + \frac{\Omega^{13}e_1 + \Omega^{23}e_2}{\Omega^{33}}\right)^2\right\}$$

$$\begin{aligned}
& \times \exp\left[-\frac{1}{2} \left\{ \left(\frac{\Omega^{11}\Omega^{33} - (\Omega^{13})^2}{\Omega^{33}} \right) e_1^2 + 2 \left(\frac{\Omega^{13}\Omega^{23} - \Omega^{12}\Omega^{33}}{\Omega^{33}} \right) e_1 e_2 \right. \right. \\
& \quad \left. \left. + \left(\frac{\Omega^{22}\Omega^{33} - (\Omega^{23})^2}{\Omega^{33}} \right) e_2^2 \right\} \right] \\
& = \frac{1}{(2\pi)^{1/2} (\Omega^{33})^2} \exp\left\{-\frac{\Omega^{33}}{2} \left(e_3 + \frac{\Omega^{13}e_1 + \Omega^{23}e_2}{\Omega^{33}} \right)^2\right\} \\
& \quad \times f(e_1, e_2, \rho)
\end{aligned}$$

where $\Omega^{33} = (1 - \rho^2) / |\Omega|$. Then,

$$\begin{aligned}
(A.4) \int_{-\infty}^{\infty} e_3 f(e_1, e_2, e_3) de_3 &= \frac{\sigma_{13}^{-\rho\sigma_{23}}}{1 - \rho^2} e_1 f(e_1, e_2, \rho) \\
&+ \frac{\sigma_{23}^{-\rho\sigma_{13}}}{1 - \rho^2} e_2 f(e_1, e_2, \rho)
\end{aligned}$$

Substituting (A.4) into (A.3) and using Rosenbaum's theorem (See Johnson and Kotz [1972]) gives us (A.1).

To show (A.2) to hold, note that

$$\begin{aligned}
(A.5) E(e_3^2 | e_1 \geq -x_1\beta_1, e_2 \geq -x_2\beta_2) &= \frac{1}{F(x_1\beta_1, x_2\beta_2, \rho)} \\
&\times \int_{-x_1\beta_1}^{\infty} \int_{-x_2\beta_2}^{\infty} \int_{-\infty}^{\infty} e_3^2 f(e_1, e_2, e_3) de_3 de_2 de_1
\end{aligned}$$

It can be easily shown that

$$\begin{aligned}
(A.6) \int_{-\infty}^{\infty} e_3^2 f(e_1, e_2, e_3) de_3 &= \frac{|\Omega|}{1 - \rho^2} f(e_1, e_2, \rho) \\
&+ \frac{(\rho\sigma_{23} - \sigma_{13})^2}{1 - \rho^2} e_1^2 f(e_1, e_2, \rho) \\
&+ \frac{2(\rho\sigma_{23} - \sigma_{13})(\rho\sigma_{13} - \sigma_{23})}{(1 - \rho^2)^2} e_1 e_2 f(e_1, e_2, \rho)
\end{aligned}$$

$$+ \frac{(\rho\sigma_{13}^2 - \sigma_{23})}{(1-\rho^2)^2} e_2^2 f(e_1, e_2, \rho)$$

Substituting (A.6) into (A.5), and using Rosenbaum's theorem again gives us (A.2).

(A.2) provides a consistent estimator of σ_{33} in the extended two-stage estimation method. As we see in (4.3'), $v_{3i} = e_{3i} - \sigma_{13}\mu_{1i} - \sigma_{23}\mu_{2i}$. Then using (A.2), we can show that

$$(A.7) \quad E(v_{3i}^2 | e_{1i} \geq -x_{1i}\beta_1, e_{2i} \geq -x_{2i}\beta_2) = \sigma_{33} - \sigma_{13}^2(x_{1i}\beta_1\mu_{1i} + \mu_{1i}^2 + \rho\mu_{3i}) \\ - \sigma_{23}(x_{2i}\beta_2\mu_{2i} + \mu_{2i}^2 + \rho\mu_{3i}) + 2\sigma_{13}\sigma_{23}(\mu_{3i} - \mu_{1i}\mu_{2i})$$

This shows the consistency of $\hat{\sigma}_{33}$ given in (10).

Appendix B

The extended two-stage estimators of β_3 and c are given by:

$$(B.1) \quad \alpha = \begin{pmatrix} \hat{\beta}_3 \\ \hat{c} \end{pmatrix} = \begin{pmatrix} \Sigma_i x'_{3i} x_{3i} & \Sigma_i x'_{3i} \hat{\mu}_i \\ \Sigma_i \hat{\mu}_i' x_{3i} & \Sigma_i \hat{\mu}_i' \hat{\mu}_i \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_i x'_{3i} y_{3i} \\ \Sigma_i \hat{\mu}_i' y_{3i} \end{pmatrix}$$

$$= \alpha + \begin{pmatrix} \Sigma_i x'_{3i} x_{3i} & \Sigma_i x'_{3i} \hat{\mu}_i \\ \Sigma_i \hat{\mu}_i' x_{3i} & \Sigma_i \hat{\mu}_i' \hat{\mu}_i \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_i x'_{3i} \{(\mu_i - \hat{\mu}_i) + v_{3i}\} \\ \Sigma_i \hat{\mu}_i' \{(\mu_i - \hat{\mu}_i) C + v_{3i}\} \end{pmatrix}$$

where Σ_i means summation over i from 1 to N_1 .

First of all, we need to know the asymptotic distribution of $\hat{\mu}_i$. By Taylor's expansion at the true parameter, $r = (\beta_1', \beta_2', \rho)'$,

$$\hat{\mu}_i' \approx \mu_i' + \frac{\partial \mu_i'}{\partial r} (\hat{r} - r) = \mu_i' + A_i (\hat{r} - r)$$

where A_i is defined in (9). Therefore,

$$\sqrt{N}(\hat{\mu}_i' - \mu_i') \approx A_i(\sqrt{N}(\hat{r} - r)).$$

Hence, we have

$$(B.2) \quad \sqrt{N}(\hat{\mu}_i' - \mu_i') \rightarrow N(0, A_i \Psi A_i')$$

which shows the consistency of $\hat{\mu}_i$. Note that the total number (N) of observations increases as the number (N_1) of the observed y_{3i} 's does; $N \rightarrow \infty$ as $N_1 \rightarrow \infty$. For simplicity, we assume that

$$(B.3) \quad \text{plim} \frac{N_1}{N} = k, \quad 0 < k < 1$$

(See Heckman [1979]) Using (B.2) and (B.3), we can easily

show that

$$(B.4) \quad \text{plim} \frac{1}{N_1} \begin{bmatrix} \Sigma_i x'_{3i} x_{3i} & \Sigma_i x'_{3i} \hat{\mu}_i \\ \Sigma_i \hat{\mu}'_i x_{3i} & \Sigma_i \hat{\mu}'_i \hat{\mu}_i \end{bmatrix} = \text{plim} \frac{1}{N_1} \begin{bmatrix} \Sigma_i x'_{3i} x_{3i} & \Sigma_i x'_{3i} \mu_i \\ \Sigma_i \mu'_i x_{3i} & \Sigma_i \mu'_i \mu_i \end{bmatrix} = D^{-1}$$

$$(B.5) \quad \text{plim} \frac{1}{N_1} \begin{bmatrix} \Sigma_i x'_{3i} \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \\ \Sigma_i \hat{\mu}'_i \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \end{bmatrix} = 0.$$

(B.3) and (B.4) guarantee the consistency of $\hat{\alpha}$. In order to show the asymptotic covariance of $\sqrt{N_1} (\hat{\alpha} - \alpha)$, note that

$$(B.6) \quad \frac{1}{\sqrt{N_1}} \begin{bmatrix} \Sigma_i x'_{3i} \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \\ \Sigma_i \hat{\mu}'_i \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \end{bmatrix} \cdot \frac{1}{\sqrt{N_1}} \begin{bmatrix} \Sigma_i x'_{3i} \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \\ \Sigma_i \hat{\mu}'_i \{(\mu_i - \hat{\mu}_i) c + v_{3i}\} \end{bmatrix}^{-1}$$

$$= \frac{1}{N_1} \begin{bmatrix} \Sigma_i \Sigma_j & \eta_{ij} x'_{3i} x_{3j} & \Sigma_i \Sigma_j & \eta_{ij} x'_{3i} \hat{\mu}_j \\ \Sigma_i \Sigma_j & \eta_{ij} \hat{\mu}'_i x_{3j} & \Sigma_i \Sigma_j & \eta_{ij} \hat{\mu}'_i \hat{\mu}_j \end{bmatrix}$$

where $\eta_{ij} = [(\mu_i - \hat{\mu}_i) c + v_{3i}]^2$. Now, using (B.2) and (B.3) we have

$$(B.7) \quad \text{plim} \frac{1}{N_1} \Sigma_i \Sigma_j \eta_{ij} x'_{3i} x_{3j}$$

$$= \text{plim} \frac{1}{N_1} \Sigma_i \Sigma_j c' (\mu_i - \hat{\mu}_i)' (\mu_i - \hat{\mu}_i) c x'_{3i} x_{3j}$$

$$+ \text{plim} \frac{2}{N_1} \Sigma_i \Sigma_j v_{3i} (\mu_j - \hat{\mu}_j) c x'_{3i} x_{3j}$$

$$+ \text{plim} \frac{1}{N_1} \Sigma_i \Sigma_j v_{3i} v_{3j} x'_{3i} x_{3j}$$

$$\begin{aligned}
&= \text{plim} \frac{N_1}{N} \left(\frac{\sum_i \sum_j c_{ij} A_i' \psi A_i' c x_{3i}' x_{3j}}{N_1^2} \right) \\
&\quad + \text{plim} \frac{1}{N_1} \sum_i v_{3i}^2 x_{3i}' x_{3i} \\
&= \text{plim} \frac{N_1}{N} \left(\frac{\sum_i \sum_j Q_{ij} x_{3i}' x_{3j}}{N_1^2} \right) + \text{plim} \frac{1}{N} \sum_i \pi_{ii} x_{3i}' x_{3i}
\end{aligned}$$

Here π_{ii} and Q_{ij} are given in (9). The last equality in (B.7) is derived from (A.7). The limiting values of the other components in (B.6) can be derived by similar methods. Therefore, we have

$$\begin{aligned}
\text{(B.8) } &\text{plim} \frac{1}{N_1} \begin{bmatrix} \sum_i \sum_j \eta_{ij} x_{3i}' x_{3j} & \sum_i \sum_j \eta_{ij} x_{3i}' \hat{\mu}_j \\ \sum_i \sum_j \eta_{ij} \hat{\mu}_i' x_{3j} & \sum_i \sum_j \eta_{ij} \hat{\mu}_i' \hat{\mu}_j \end{bmatrix} \\
&= \sigma_{33} \text{plim} \frac{1}{N_1} \begin{bmatrix} \sum_i \pi_{ii} x_{3i}' x_{3j} & \sum_i \pi_{ii} x_{3i}' x_{3i} \mu_i \\ \sum_i \pi_{ii} \mu_i' x_{3j} & \sum_i \pi_{ii} \mu_i' \mu_i \end{bmatrix} \\
&\quad + \text{plim} \left(\frac{N_1}{N} \right) \frac{1}{N_1^2} \begin{bmatrix} \sum_i \sum_j Q_{ij} x_{3i}' x_{3j} & \sum_i \sum_j Q_{ij} x_{3i}' \mu_j \\ \sum_i \sum_j Q_{ij} \mu_i' x_{3j} & \sum_i \sum_j Q_{ij} \mu_i' \mu_j \end{bmatrix}
\end{aligned}$$

Substituting (B.4) and (B.8) gives us

$$\text{(B.9) } \sqrt{N}(\hat{\alpha} - \alpha) \rightarrow N(0, \text{DBD}')$$

Appendix C

Here, we will derive the LM test statistic for no selection bias in Model II. To construct the LM test statistic, we need the restricted MLE estimators. Let $a_1 = \sigma_{13}/\sigma_{33}$; $a_2 = \sigma_{23}/\sigma_{33}$.

Note that $\sigma_{13} = \sigma_{23} = 0$ if and only if $a_1 = a_2 = 0$. Therefore, the null hypothesis of no sample selection bias is that $a_1 = a_2 = 0$. We can rewrite the log-likelihood function for Model II as

$$\begin{aligned}
 (C.1) \quad \ln L^{II}(\beta_1, \beta_2, \rho, \beta_3, a_1, a_2, \sigma_{33}) \\
 = \sum_{i=1}^N [Y_{1i}Y_{2i} \{ -\frac{1}{2} \ln \sigma_{33} - \frac{1}{2\sigma_{33}} (Y_{3i} - x_{3i}\beta_3) \\
 + \ln \left[F \left[\frac{x_{1i}\beta_1 + a_1(Y_{3i} - x_{3i}\beta_3)}{(1-a_1^2\sigma_{33})^{1/2}}, \frac{x_{2i}\beta_2 + a_2(Y_{3i} - x_{3i}\beta_3)}{(1-a_2^2\sigma_{33})^{1/2}}, \right. \right. \\
 \left. \left. \frac{\rho - \sigma_{33}a_1a_2}{(1-a_1^2\sigma_{33})^{1/2}(1-a_2^2\sigma_{33})^{1/2}} \right] \right] \\
 - \ln F(x_{1i}\beta_1, x_{2i}\beta_2, \rho) \} + \ln L^C(\beta_1, \beta_2, \rho)
 \end{aligned}$$

where $\ln L^C(\beta_1, \beta_2, \rho)$ is given in (6). Let the restricted MLE estimates of β_1 , β_2 , ρ , β_3 and σ_{33} be $\bar{\beta}_1$, $\bar{\beta}_2$, $\bar{\rho}$, $\bar{\beta}_3$, and $\bar{\sigma}_{33}$. Under the null hypothesis, (C.1) is reduced to

$$\begin{aligned}
 (C.2) \quad \ln L^{II}(\beta_1, \beta_2, \beta_3, \rho, 0, 0, \sigma_{33}) \\
 = \sum_{i=1}^N [Y_{1i}Y_{2i} \{ -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_{33} - \frac{1}{2\sigma_{33}} (Y_{3i} - x_{3i}\beta_3)^2 \}] \\
 + \ln L^C(\beta_1, \beta_2, \rho)
 \end{aligned}$$

This shows that $\bar{\beta}_3$, $\bar{\sigma}_{33}$ can be derived from the usual OLS

procedure for the observed y_{3i} 's, and that $\bar{\beta}_1, \bar{\beta}_2, \bar{\rho}$ are just MLE estimators derived by maximizing $\ln L^C(\beta_1, \beta_2, \rho)$. Therefore, $\bar{\beta}_1, \bar{\beta}_2$, and $\bar{\rho}$ are exactly identical to the estimates, $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\rho}$ used for the 2-stage estimation.

Let $\theta = (\beta_1' \beta_2' \rho \beta_3' a_1 a_2 \sigma_{33})'$. Then the restricted MLE estimator of θ is given by:

$$\bar{\theta} = (\bar{\beta}_1', \bar{\beta}_2', \bar{\rho}, \bar{\beta}_3', 0, 0, \bar{\sigma}_{33})' = (\hat{\beta}_1', \hat{\beta}_2', \hat{\rho}, \bar{\beta}_3', 0, 0, \bar{\sigma}_{33})'$$

where $\bar{\sigma}_{33} = \sum_i (y_{3i} - x_{3i}\bar{\beta}_3)^2 / N_1$. Then, we can show

(C.3)

$$\frac{\partial \ln L^{II}}{\partial \theta} \Big|_{\bar{\theta}} = (0, 0, 0, 0, \sum_{i=1}^{N_1} \hat{\mu}_{1i}(y_{3i} - x_{3i}\bar{\beta}_3), \sum_{i=1}^{N_1} \hat{\mu}_{2i}(y_{3i} - x_{3i}\bar{\beta}_3), 0)'$$

Some tedious operations generate the information matrix evaluated at $\bar{\theta}$:

$$(C.4) \quad I(\theta) = E\left(-\frac{\partial^2 \ln L^{II}}{\partial \theta \partial \theta} \Big|_{\bar{\theta}}\right)$$

$$= \begin{bmatrix} E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_1 \partial \beta_1} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_1 \partial \beta_2} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_1 \partial \rho} \Big|_{\bar{\theta}}\right) \\ E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_2 \partial \beta_1} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_2 \partial \beta_2} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \beta_2 \partial \rho} \Big|_{\bar{\theta}}\right) \\ E\left(-\frac{\partial^2 \ln L^C}{\partial \rho \partial \beta_1} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \rho \partial \beta_2} \Big|_{\bar{\theta}}\right) & E\left(-\frac{\partial^2 \ln L^C}{\partial \rho^2} \Big|_{\bar{\theta}}\right) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
(1/\bar{\sigma}_{33}) \sum_{i=1}^N \hat{F}_i x'_{3i} x_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x'_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x'_{3i} & 0 \\
\sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i}^2 & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} & 0 \\
\sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i}^2 & 0 \\
0 & 0 & 0 & (\bar{\sigma}_{33}/2) \sum_{i=1}^N \hat{F}_i
\end{array}
\left. \vphantom{\begin{array}{cccc} 0 & 0 & 0 & 0 \end{array}} \right]$$

Now, using (C.3) and (C.4), we have the LM statistic:

$$\begin{aligned}
\text{(C.5) } LM &= \left(\frac{\partial \ln L^{II}}{\partial \theta'} \bigg|_{\bar{\theta}} \right)' I(\bar{\theta})^{-1} \left(\frac{\partial \ln L^{II}}{\partial \theta} \bigg|_{\bar{\theta}} \right) \\
&= [0, \sum_{i=1}^{N_1} \hat{\mu}_{1i} (y_{3i} - x_{3i} \bar{\beta}_3), \sum_{i=1}^{N_1} \hat{\mu}_{2i} (y_{3i} - x_{3i} \bar{\beta}_3)] \\
&\quad \cdot \begin{bmatrix} (1/\bar{\sigma}_{33}) \sum_{i=1}^N \hat{F}_i x'_{3i} x_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x'_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x'_{3i} \\ \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i}^2 & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} \\ \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i}^2 \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} 0 \\ \sum_{i=1}^{N_1} \hat{\mu}_{1i} (y_{3i} - x_{3i} \bar{\beta}) \\ \sum_{i=1}^{N_1} \hat{\mu}_{2i} (y_{3i} - x_{3i} \bar{\beta}) \end{bmatrix}
\end{aligned}$$

By theorem 4.2.3 in Amemiya [1985] (p. 118),

$$(C.6) \quad \text{plim} \frac{1}{N} \sum_{i=1}^N (y_{1i} y_{2i} - F_{2i}) H_i = 0$$

where H_i is a nonstochastic variable. Since $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\rho}$ are consistent,

$$(C.7) \quad \text{plim} F(x_{1i}\beta_1, x_{2i}\beta_2, \rho) = F(x_{1i}\beta_1, x_{2i}\beta_2, \rho)$$

Using (C.6) and (C.7), we can show that

$$(C.8) \quad \text{plim} \frac{1}{N} \begin{bmatrix} (1/\bar{\sigma}_{33}) \sum_{i=1}^N \hat{F}_i x'_{3i} x_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x'_{3i} & \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x'_{3i} \\ \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i}^2 & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} \\ \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{1i} \hat{\mu}_{2i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{F}_i \hat{\mu}_{2i}^2 \end{bmatrix}$$

$$= \text{plim} \frac{1}{N} \begin{bmatrix} (1/\bar{\sigma}_{33}) \sum_{i=1}^N y_{1i} y_{2i} x'_{3i} x_{3i} & \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{1i} x'_{3i} & \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{2i} x'_{3i} \\ \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{1i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{1i}^2 & \bar{\sigma}_{33} \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{1i} \hat{\mu}_{2i} \\ \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{2i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{1i} \hat{\mu}_{2i} & \bar{\sigma}_{33} \sum_{i=1}^N y_{1i} y_{2i} \hat{\mu}_{2i}^2 \end{bmatrix}$$

$$= \text{plim } \frac{1}{N} \begin{bmatrix} (1/\bar{\sigma}_{33}^{-1}) \sum_{i=1}^N x_{3i}' x_{3i} & \sum_{i=1}^N \hat{u}_{1i} x_{3i}' & \sum_{i=1}^N \hat{\mu}_{2i} x_{3i}' \\ \sum_{i=1}^N \hat{\mu}_{1i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{\mu}_{1i}^2 & \bar{\sigma}_{33} \sum_{i=1}^N \hat{\mu}_{1i} \hat{\mu}_{2i} \\ \sum_{i=1}^N \hat{\mu}_{2i} x_{3i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{\mu}_{1i} \hat{\mu}_{2i} & \bar{\sigma}_{33} \sum_{i=1}^N \hat{\mu}_{2i}^2 \end{bmatrix}$$

Using (C.8) and matrix notation defined in Section II, we have:

$$\begin{aligned} \text{(C.9) } LM &\approx (Y_3 - X_3 \bar{\beta}_3)' \hat{\mu} (\bar{\sigma}_{33} \hat{\mu}' \hat{\mu} - \bar{\sigma}_{33} \hat{\mu}' X_3 (X_3' X_3)^{-1} X_3 \hat{\mu})^{-1} \hat{\mu}' (Y_3 - X_3 \bar{\beta}_3) \\ &= \frac{\hat{c}' \hat{\mu}' M_X \hat{\mu} \hat{c}}{\bar{\sigma}_{33}} \rightarrow \chi^2(2) \end{aligned}$$

using the facts;

$$\bar{\beta}_3 = (X_3' X_3)^{-1} X_3' Y_3$$

$$M_X (Y_3 - X_3 \bar{\beta}_3) = Y_3 - X_3 \bar{\beta}_3$$

$$\hat{c} = (\hat{\mu}' M_X \hat{\mu})^{-1} \hat{\mu}' M_X Y_3$$

Now we can see that the LM statistic is identical to the F statistic given in Section II except for the difference between $SSE_0/(N_1 - k_3)$ and $\bar{\sigma}_{33}$. Note that

$$F = \frac{\hat{c}' \hat{\mu}' M_X \hat{\mu} \hat{c} / 2}{SSE_0 / (N_1 - k_3)} \rightarrow \chi^2(2) / 2$$

and $SSE_0/(N_1 - k_3)$ and $\bar{\sigma}_{33}$ have no substantial asymptotic difference. Therefore, the F-test is identical to the LM test, and has good asymptotic power properties.

Appendix D

Let $(e_1, e_2, e_3)'$ have the trivariate normal distribution with zero mean and covariance matrix

$$\Omega = \begin{bmatrix} 1 & \rho & \sigma_{13} \\ \rho & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

Then, the conditional distribution of e_3 given $e_1 > -x$ and $e_2 > -y$ is:

$$\begin{aligned} (D.1) \quad f(e_3 | e_1 \geq -x, e_2 \geq -y) &= \frac{1}{(2\pi)^{1/2} (\sigma_{33})^{1/2}} \exp \left(-\frac{1}{2\sigma_{33}} e_3^2 \right) \\ &\cdot F \left[\frac{x + (\sigma_{13}/\sigma_{33}) e_3}{(1 - \sigma_{13}^2/\sigma_{33})^{1/2}}, \frac{y + (\sigma_{23}/\sigma_{33}) e_3}{(1 - \sigma_{23}^2/\sigma_{33})^{1/2}}, \right. \\ &\quad \left. \frac{\rho - \sigma_{13}\sigma_{23}/\sigma_{33}}{(1 - \sigma_{13}^2/\sigma_{33})^{1/2} (1 - \sigma_{23}^2/\sigma_{33})^{1/2}} \right] \\ &\div F(x, y, \rho) \end{aligned}$$

Proof. It can be easily shown that

$$\begin{aligned} (D.2) \quad F(e_1, e_2 | e_3) &= \frac{1}{2\pi (|\Omega|/\sigma_{33})^{1/2}} \exp \left[-\frac{1}{2|\Omega|} \{ (\sigma_{33} - \sigma_{23}) \right. \\ &\quad \times \exp \left[-\frac{1}{2|\Omega|} \{ (\sigma_{33} - \sigma_{23}^2) (e_1 - \frac{\sigma_{13}}{\sigma_{33}} e_3)^2 \right. \\ &\quad \left. \left. - 2(\rho\sigma_{33} - \sigma_{23}\sigma_{13}) (e_1 - \frac{\sigma_{13}}{\sigma_{33}} e_3)^2 \right. \right. \\ &\quad \left. \left. + (\sigma_{33} - \sigma_{13}^2) (e_2 - \frac{\sigma_{23}}{\sigma_{33}} e_3)^2 \} \} \right] \end{aligned}$$

The conditional distribution of e_3 given $e_1 > -x$ and $e_2 > -y$

is described by

$$\begin{aligned}
 (D.3) \quad F(e_3 | e_1 > -x, e_2 > -y) &= \int_{-x}^{\infty} \int_{-y}^{\infty} \frac{f(e_1, e_2, \rho_3)}{F(x, y, \rho)} de_2 de_1 \\
 &= \frac{f(e_3)}{F(x, y, \rho)} \int_{-x}^{\infty} \int_{-y}^{\infty} f(e_1, e_2 | e_3) de_2 de_1
 \end{aligned}$$

where

$$\begin{aligned}
 f(e_3) &= \frac{1}{(2\pi)^{1/2} (\sigma_{33})^{1/2}} \exp \left(-\frac{e_3^2}{2\sigma_{33}} \right) \\
 (D.4) \quad t_1 &= \frac{e_1 - (\sigma_{13}/\sigma_{33})e_3}{(1 - \sigma_{13}^2/\sigma_{33})^{1/2}}; \quad t_2 = \frac{e_2 - (\sigma_{23}/\sigma_{33})e_3}{(1 - \sigma_{23}^2/\sigma_{33})^{1/2}}
 \end{aligned}$$

Then, the Jacobian of the transformation is

$$(D.5) \quad \left| \frac{\partial(e_1, e_2)}{\partial(t_1, t_2)} \right| = \frac{(\sigma_{33} - \sigma_{13}^2)^{1/2} (\sigma_{33} - \sigma_{23}^2)^{1/2}}{\sigma_{33}}$$

Using (D.4) and (D.5), we have

$$\begin{aligned}
 (D.6) \quad &\int_{-x}^{\infty} \int_{-y}^{\infty} f(e_1, e_2 | e_3) de_2 de_1 \\
 &= \int_h^{\infty} \int_k^{\infty} g(t_1, t_2) dt_2 dt_1
 \end{aligned}$$

where

$$h = -(x + (\sigma_{13}/\sigma_{33})e_3) / (1 - \sigma_{13}^2/\sigma_{33})^{1/2}$$

$$k = -(y + (\sigma_{23}/\sigma_{33})e_3) / (1 - \sigma_{23}^2/\sigma_{33})^{1/2}$$

$$l = (\rho\sigma_{33} - \sigma_{13}\sigma_{23}) / \{(\sigma_{33} - \sigma_{13}^2)^{1/2} (\sigma_{33} - \sigma_{23}^2)^{1/2}\}$$

$$\begin{aligned}
g(t_1, t_2) = & \frac{(\sigma_{33} - \sigma_{13}^2)^{1/2} (\sigma_{33} - \sigma_{23}^2)^{1/2}}{2\pi(|\Sigma|\sigma_{33})^{1/2}} \\
& \cdot \exp\left[- \frac{(\sigma_{33} - \sigma_{13}^2)(\sigma_{33} - \sigma_{23}^2)}{2|\Sigma|\sigma_{33}} \right. \\
& \left. \cdot \left\{ t_1^2 + t_2^2 - \frac{2(\rho\sigma_{33} - \sigma_{13}\sigma_{23})}{(\sigma_{33} - \sigma_{13}^2)^{1/2}(\sigma_{33} - \sigma_{23}^2)^{1/2}} t_1 t_2 \right\} \right]
\end{aligned}$$

Note that

$$1 - \ell^2 = \frac{|\Sigma|\sigma_{33}}{(\sigma_{33} - \sigma_{13}^2)(\sigma_{33} - \sigma_{23}^2)}$$

This shows that $g(t_1, t_2)$ is the standard bivariate normal distribution function with the correlation coefficient, ℓ .

Therefore, we have

$$(D.7) \quad \int_{-x}^{\infty} \int_{-y}^{\infty} f(e_1, e_2 | e_3) de_2 de_1 = F(h, k, \ell)$$

Substituting (D.7) into (D.3) gives us (D.1).

REFERENCES

REFERENCES

- Abowd, J.M., and Farber, H.S. (1982), "Job Queues and the Union Status of Workers", Industrial and Labor Relations Review, Vol. 35, pp 354 - 368.
- Amemiya, T (1973), "Regression Analysis When the Dependent Variable Is Truncated Normal", Econometrica, Vol. 41, pp 997 - 1016.
- Amemiya, T (1985), Advanced Econometrics, Harvard Press.
- Blundell, R., Ham, J., and Meghir, C. (1987), "Unemployment and Female Labor Supply", The Economic Journal, Vol. 97 (Conference Papers), pp 44 - 64.
- Farber, H.S. (1983), "Worker Preference for Union Representation," Research in Labor Economics, Supplement 2, pp 171 - 205.
- Fishe, R., Trost, R.P., and Lurie, P.M. (1981), "Labor Force Earnings and College Choice of Young Women: An Examination of Selectivity Bias and Comparative Advantage", Economics of Education Review, Vol. 1, pp 169 - 191.
- Flinn, C. J., and Heckman, J.J. (1983), "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?", Journal of Labor Economics, Vol. 1, pp 169 - 191.
- Ham, J.C. (1982), "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment", Review of Economic studies, Vol. 49, pp 335 - 354.
- Heckman, J. (1974), "Shadow Prices, Market wages, and Labor Supply", Econometrica, Vol. 42, pp 670 - 694.
- Heckman, J. (1979), "Sample Selection Bias As a Specification Error", Econometrica, Vol. 47, pp 153 - 161.
- Johnson, N.L., and Kotz, S. (1972), Distributions in Statistics: Continuous Multivariate Distributions, Vol. 4, John Wiley and Sons.
- Lin, T. (1982), Some Applications of the Lagrangean Multiplier Test in Econometrics, Dissertation for the Ph.D degree, Michigan State University.
- Maddala, G.S. (1987), Limited-dependent And Qualitative Variables in Econometrics, Econometric Society Monograph, No. 3, Cambridge University Press.

- Melino, A. (1982), "Testing for Sample Selection Bias", Review of Economic Studies, Vol. XLIX, pp 151 - 153.
- Meng, C., and Schmidt, P. (1985), "On the Cost of Partial Observability in the Bivariate Probit Model", International Economic Review, Vol. 26, pp 71 - 85.
- Poirier, D.J. (1980), "Partial Observability in Bivariate Probit Models", Journal of Econometrics, Vol. 12, pp 209 - 217.

Chapter 3

Efficient Estimation of Models for Dynamic Panel Data

I. Introduction

This paper considers a dynamic model with panel data which include a large number of cross-section observations, but only over a short period of time. A typical problem in using panel data is that the error terms in the model contain unobservable and time-invariant individual effects. To allow for these effects, "random effects" models are widely used in the literature on dynamic panel data. In these models, the individual effects are treated as being generated from an independently identically distributed (iid) stochastic process. This paper develops a generalized- method-of-moments (GMM) estimator for the dynamic model with random effects which is efficient under general circumstances.

In the case of the static model, the simple fixed effects (within) treatment generates a consistent estimator. There are also a number of studies which develop efficient estimation methods for the static model with random effects. When no explanatory variables are correlated with the individual effects, the generalized least square (GLS) estimator is consistent and efficient in finite sample; see Hsiao [1986]. When some explanatory variables are correlated with the individual effects, we can efficiently estimate the model using some available instrument variables; see Hausman

and Taylor [1981], Amemiya and MaCurdy [1986], and Breusch, Mizon, and Schmidt [1989].

Several problems arise in the dynamic model that do not arise in the static model. First, the conventional within estimator is inconsistent unless there are a large number of time-series observations; see Hsiao [1986]. Second, even though the maximum-likelihood (ML) method is available, the form of the ML estimator depends crucially on assumptions about the initial observations and the distribution of the individual effect; see Anderson and Hsiao [1981], or Hsiao [1982]. To avoid these problems, Anderson and Hsiao [1981], Holtz-Eakin [1988], and Arellano and Bond [1988] investigate instrumental variables estimation techniques. (From now on I call their methods the conventional instrumental-variable (IV) methods.) To get a consistent estimator, they first-difference the original equation to eliminate the individual effects; and then they use lagged dependent variables as instruments. These instruments are legitimate in the usual sense that they are uncorrelated with the differenced error terms.

In the framework of Hansen's [1982] GMM, the conventional IV estimators can be regarded as GMM estimators which use some available linear orthogonality conditions. The GMM estimators are efficient in general circumstances, if all known is that the data-generating process satisfies certain moment restrictions; see Chamberlain [1987]. The GMM method may be preferred to the ML methods in the dynamic model using panel

data, because the GMM estimators do not rely on assumptions about the initial observations and the distribution of the individual effects. However, the conventional IV estimators are not efficient in the sense that they fail to use all the available moment conditions. This is due to their lack of a systematic treatment in counting the number of available restrictions. Furthermore, those IV estimators could be inconsistent if we relax some behavioral assumptions about the error terms. For example, lagged dependent variables are generally not legitimate instruments if the error terms are autocorrelated.

The main goal of this paper is to offer a systematic analysis which counts properly all the available moment conditions under given assumptions. Under alternative sets of assumptions, I demonstrate how many moment conditions there may be, and I show how to write them in a convenient form. Under the usual assumptions, we can find some linear and nonlinear orthogonality conditions, which the conventional IV approaches do not exploit. I then derive the GMM estimator, and a linearized GMM estimator that is equally asymptotically efficient. The GMM estimator based on all the available moment conditions must be more efficient than other GMM estimators based on only a subset of moment conditions. In this sense, the GMM estimator presented in this paper could be said to be efficient when the distributions of the initial observations and of the individual effects are not known.

The plan of this paper is as follows. Sections II, III,

and IV consider a simple dynamic model which includes only a one-period lagged dependent variable as the explanatory variable. Section II briefly summarizes the conventional IV approaches and demonstrates why they miss some available moment restrictions. Section III shows the proper way to derive all the legitimate moment conditions under given assumptions. Section IV investigates the estimation procedure and the asymptotic performance of our GMM estimator under the usual assumptions. Section V extends our approach to the dynamic model which includes exogenous variables. Section VI gives some conclusion.

II. Conventional IV Methods

To explain the conventional IV methods as simply as possible, I consider the following simple dynamic model:

$$\begin{aligned}
 \text{(II.1-1)} \quad y_{it} &= \delta y_{i,t-1} + \alpha_i + \epsilon_{it} \\
 &\quad (\text{for } i = 1, 2, \dots, N; t = 1, 2, \dots, T) \\
 &= \delta y_{i,t-1} + u_{it}
 \end{aligned}$$

where $u_{it} = \alpha_i + \epsilon_{it}$. The subscript i denotes the i th cross-section unit, and t designates time periods. Here y is the dependent variable, α is the individual effect, and ϵ is the error term. Let $y_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$; $y_{i,-1} = (y_{i0}, y_{i1}, \dots, y_{i,T-1})'$, $u_i = (u_{i1}, u_{i2}, \dots, u_{iT})'$. Then, we can rewrite (II.1-1) as

$$\begin{aligned}
 \text{(II.1-2)} \quad y_i &= \delta y_{i,-1} + u_i \\
 i &= 1, 2, \dots, N
 \end{aligned}$$

Let $y = (y_1', y_2', \dots, y_N')'$; $y_{-1} = (y_{1,-1}', y_{2,-1}', \dots, y_{N,-1}')'$;

$u=(u_1', u_2', \dots, u_N')'$. Then (II.1-2) becomes

$$(II.1-3) \quad y = \delta y_{-1} + u.$$

I assume that the α 's and ϵ 's have zero means. More assumptions about them will be made shortly.

For mathematical convenience and future use, I also define some notation:

$$P_A = A(A'A)^{-1}A'$$

$$M_A = I_m - P_A$$

$$P = (1/T)e_T e_T'$$

$$Q = I_T - P$$

$$P_v = I_N \otimes P$$

$$Q_v = I_N \otimes Q = I_{NT} - P_v$$

where e_T is a T -dimensional vector of ones, and A is any $m \times k$ matrix.

In the conventional literature on the dynamic panel data model, there are four common assumptions about y_0 , α , and ϵ 's (for simplicity, the subscript "i" is suppressed):

(SA.1) ϵ 's are independent of y_0 ; i.e., $E(y_0 \epsilon_t) = 0$ for any t .

(SA.2) ϵ 's are independent of α ; i.e., $E(\alpha \epsilon_t) = 0$ for any t .

(SA.3) ϵ 's are homoskedastic; i.e., $E(\epsilon_t^2) = \sigma_\epsilon^2$ for any t .

(SA.4) ϵ 's are mutually independent; i.e., $E(\epsilon_s \epsilon_t) = 0$ for any $t \neq s$.

The conventional IV approaches first-difference (II.1):

$$(II.2) \quad y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (u_{it} - u_{i,t-1})$$

Since $u_{it} - u_{i,t-1}$ ($= \epsilon_{it} - \epsilon_{i,t-1}$) does not include α_i , some lagged y_{it} 's can be used as instruments for the estimation of δ . For example, $y_{i,t-2} - y_{i,t-3}$ is a legitimate instrumental variable in

the sense that it is uncorrelated with $u_{it}-u_{i,t-1}$, but correlated with $y_{i,t-1}-y_{i,t-2}$; see Anderson and Hsiao [1981].

Arellano and Bond [1988] and Holtz-Eakin [1988] find and use all of the moment conditions based on lagged y 's being uncorrelated with $u_{it}-u_{i,t-1}$. Consider the $(T-1)$ first-differenced equations separately:

$$(II.3) \quad y_{i2}-y_{i1} = \delta(y_{i1}-y_{i0}) + (u_{i2}-u_{i1})$$

$$y_{i3}-y_{i2} = \delta(y_{i2}-y_{i1}) + (u_{i3}-u_{i2})$$

:

$$y_{iT}-y_{i,T-1} = \delta(y_{i,T-1}-y_{i,T-2}) + (u_{iT}-u_{i,T-1})$$

The system (II.3) could be regarded as a simultaneous system of equations with the cross-equation restriction that the coefficients are the same everywhere. Arellano and Bond's method is akin to three stage least squares (3SLS) with different instruments for the different equations; $y_{i0}, y_{i1}, \dots, y_{i,j-1}$ for the j th equation of (II.3). This approach is based on the following $(\frac{1}{2})T(T-1)$ orthogonality conditions:

$$(II.4) \quad E(y_{it}(u_{i,s+1}-u_{is})) = 0,$$

$$t = 1, 2, \dots, T-2; \quad t < s \leq T-1$$

which hold under the assumptions given in (SA).

Even though Anderson and Hsiao [1981], Holtz-Eakin [1988], and Arellano and Bond [1988] adopt somewhat different instrumental variable treatments, fundamentally all of their methods are based on the conditions given in (II.4). To clarify this point, let us try a slightly different approach. Define

$A_i =$

$$\begin{bmatrix} -y_{i0} & 0 & & & : & 0 & 0 & & : & : & 0 \\ y_{i0} & -y_{i0} & & & : & -y_{i1} & 0 & & : & : & 0 \\ 0 & y_{i0} & & & : & y_{i1} & -y_{i1} & & : & : & 0 \\ 0 & 0 & . & & : & 0 & y_{i1} & & : & : & 0 \\ . & . & . & & : & . & . & . & : & : & . \\ . & . & . & . & : & . & . & . & : & : & . \\ . & . & . & . & : & . & . & . & : & : & . \\ 0 & 0 & & -y_{i0} & : & 0 & 0 & & -y_{i1} & : & -y_{i,T-2} \\ 0 & 0 & & y_{i0} & : & 0 & 0 & & y_{i1} & : & y_{i,T-1} \end{bmatrix}$$

where A_i is a $T \times (\frac{1}{2})T(T-1)$ matrix. Also, define

$$A = (A_1', A_2', \dots, A_N')'$$

Then, (II.4) can be compactly expressed by

$$(II.5) \quad E(A_i' u_i) = 0 \text{ or}$$

$$E(A'u) = 0$$

One merit of (II.5) is that we do not have to first-difference the equation given in (II.1). That is, we can directly apply the instrumental-variable treatment to the equation in levels. All the conventional IV approaches can be interpreted as using some if not all orthogonality conditions which are just linear combinations of those in (II.5). Note that the instruments in A are legitimate because $\text{plim}(1/N)A'u=0$ and $\text{plim}(1/N)A'y_{-1} \neq 0$. It can be easily shown that Anderson and Hsiao use as instrumental variable a linear combination of only some columns of A , while Arellano and Bond use A itself. The IV estimator which uses A for (II.1-3) takes the following form:

$$(II.6) \quad \hat{\delta}_A = (y_{-1}' P_A y_{-1})^{-1} y_{-1}' P_A y$$

It can also be shown that this estimator is asymptotically identical to the GMM estimator based on (II.5) with

assumptions in (SA). (See APPENDIX A.) The asymptotic distribution of $\hat{\delta}_A$ is given by:

$$(II.7) \sqrt{N} (\hat{\delta}_A - \delta) \rightarrow N(0, \sigma_\epsilon^2 \text{plim}[(1/N)Y_{-1}'P_A Y_{-1}]^{-1})$$

If all the available orthogonality conditions are those in (II.5), $\hat{\delta}_A$ must be efficient among the class of estimators which can be derived using this set of information.

However, the assumptions in (SA) imply more moment restrictions than those in (II.5). For an example, consider the following second-differenced equation:

$$(II.8) Y_{i3} - Y_{i1} = \delta(Y_{i2} - Y_{i0}) + (u_{i3} - u_{i1})$$

Clearly, under the assumptions in (SA),

$$(II.9-1) E[u_{i2}(u_{i3} - u_{i1})] = 0, \text{ and}$$

$$(II.9-2) E[u_{i2}(Y_{i2} - Y_{i1})] \neq 0.$$

Therefore, u_{i2} can be regarded as a nonlinear instrument for equation (II.8) in the sense of Amemiya [1974], and this restriction (II.9-1) is not implied by (II.5). This example implies that (II.5) does not incorporate all the available moment conditions enforceable under the assumptions in (SA), and therefore that $\hat{\delta}_A$ is not efficient. In the next section I will categorize all the restrictions implied by (SA), and also show how we can relax some assumptions.

III. Derivation of Moment Conditions

In this section, I demonstrate an appropriate way to derive all the available moment restrictions which can be exploited from the usual set of assumptions given in (SA), and I also apply this method to several cases in which some

assumptions are relaxed. To do so, first define the covariance matrix of y_0 , α , and the ϵ 's¹:

$$(III.1) \quad \Sigma = \text{Cov} \begin{bmatrix} y_0 \\ \alpha \\ \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} & \sigma_{01} & \sigma_{02} & \cdots & \sigma_{0T} \\ & \sigma_\alpha^2 & \sigma_{\alpha 1} & \sigma_{\alpha 2} & \cdots & \sigma_{\alpha T} \\ & & \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} \\ & & & \sigma_{22} & \cdots & \sigma_{2T} \\ & & & & \ddots & \vdots \\ & & & & & \sigma_{TT} \end{bmatrix}$$

where i is suppressed. Basically, any assumption which may be imposed on the dynamic model can be expressed as a restriction on Σ . Under the usual assumptions in (SA), Σ takes the following form:

$$(III.2) \quad \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} & 0 & 0 & \cdots & 0 \\ & \sigma_\epsilon^2 & 0 & 0 & \cdots & 0 \\ & & \sigma_\epsilon^2 & 0 & \cdots & 0 \\ & & & \ddots & \ddots & \vdots \\ & & & & \sigma_\epsilon^2 \end{bmatrix}$$

The vector of which Σ is the covariance matrix is not observable. The vector of observables, meaning things that can be written in terms of data and parameters, is $(y_0, u_1, u_2, \cdots, u_T)'$, which has the following covariance matrix:

$$(III.3) \quad \Omega = \text{Cov} \begin{bmatrix} y_0 \\ u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \cdots & \Omega_{0T} \\ & \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1T} \\ & & \Omega_{22} & \cdots & \Omega_{2T} \\ & & & \ddots & \vdots \\ & & & & \Omega_{TT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} + \sigma_{01} & \sigma_{0\alpha} + \sigma_{02} & \cdots & \sigma_{0\alpha} + \sigma_{0T} \\ & \sigma_{\alpha}^2 + \sigma_{11} + 2\sigma_{\alpha 1} & \sigma_{\alpha}^2 + \sigma_{\alpha 1} + \sigma_{\alpha 2} + \sigma_{12} & \cdots & \sigma_{\alpha}^2 + \sigma_{\alpha 1} + \sigma_{\alpha T} + \sigma_{1T} \\ & & \sigma_{\alpha}^2 + \sigma_{22} + 2\sigma_{\alpha 2} & \cdots & \sigma_{\alpha}^2 + \sigma_{\alpha 1} + \sigma_{\alpha T} + \sigma_{2T} \\ & & & \ddots & \vdots \\ & & & & \sigma_{\alpha}^2 + \sigma_{TT} + 2\sigma_{\alpha T} \end{bmatrix}$$

By comparing Ω and Σ , we can easily see that the form of Ω depends on that of Σ , because each element of Ω is a linear combination of the elements of Σ . Under the usual assumptions (SA),

$$(III.4) \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \cdots & \Omega_{0T} \\ & \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1T} \\ & & \Omega_{22} & \cdots & \Omega_{2T} \\ & & & \ddots & \vdots \\ & & & & \Omega_{TT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} & \sigma_{0\alpha} & \cdots & \sigma_{0\alpha} \\ & \sigma_{\alpha}^2 + \sigma_{\epsilon}^2 & \sigma_{\alpha}^2 & \cdots & \sigma_{\alpha}^2 \\ & & \sigma_{\alpha}^2 + \sigma_{\epsilon}^2 & \cdots & \sigma_{\alpha}^2 \\ & & & \ddots & \vdots \\ & & & & \sigma_{\alpha}^2 + \sigma_{\epsilon}^2 \end{bmatrix}$$

An investigation of the elements in (III.4) provides us with three types of moment restrictions;

(III.5-1) Type I restrictions:

$$\Omega_{01} = \Omega_{02} = \Omega_{03} = \cdots = \Omega_{0T}$$

(III.5-2) Type II restrictions:

$$\Omega_{11} = \Omega_{12} = \Omega_{13} = \cdots = \Omega_{1T}$$

(III.5-3) Type III restrictions:

$$\begin{aligned}
\Omega_{12} &= \Omega_{13} = \Omega_{14} = \cdots = \Omega_{1T} \\
&= \Omega_{23} = \Omega_{24} = \cdots = \Omega_{2T} \\
&: \\
&= \Omega_{T-2,T-1}
\end{aligned}$$

(III.5-1) implies (T-1) restrictions; (III.5-2), (T-1); (III.5-3), $(\frac{1}{2})T(T-1)-1$. Therefore, there are in total $(\frac{1}{2})T(T-1)+(2T-3)$ available moment conditions, which are more than those used in the conventional IV approaches. Specifically, we have $(2T-3)$ extra moment conditions.

Furthermore, the conditions described in (III.5) can be obtained even under somewhat relaxed assumptions. First of all, consider Type I conditions. By observing (III.3), we can see that Type I restrictions hold as long as y_0 and ϵ_t have the same covariance for any t . The stochastic independence between y_0 and the ϵ 's is not required. Now, consider Type II conditions. These require that $\sigma_{ss} + \sigma_{\alpha}^2 + 2\sigma_{\alpha s} = \sigma_{tt} + \sigma_{\alpha}^2 + 2\sigma_{\alpha t}$ for any s and t . For this, all we need is the homoskedasticity of the ϵ 's and the equicovariance of α with the ϵ 's. Finally, consider Type III restrictions. These require $\sigma_{st} + \sigma_{\alpha t} + \sigma_{\alpha s}$ to be the same for any t and s . Obviously, the equicovariance of the ϵ 's and the equicovariance of α with the ϵ 's can justify those restrictions. In short, all the moment conditions in (III.5) are present under the following modified assumptions (MA):

(MA-1) The covariance of y_0 and ϵ_t is the same for any t .

(MA-2) The covariance of α and ϵ_t is the same for any t .

(MA-3) ϵ 's are homoskedastic.

(MA-4) The covariance of ϵ_s and ϵ_t is the same for any s and t .

A comparison of (SA) with (MA) shows that the assumptions in (SA) except the homoskedasticity of the ϵ 's are stronger than necessary.

To derive explicit forms of the moment conditions which are convenient for GMM estimation, let us rearrange the conditions in (III.5) as follows:

$$(III.6-1) \quad \Omega_{01} = \Omega_{02} = \Omega_{03} = \cdots = \Omega_{0,T-1} = \Omega_{0T}$$

$$\Omega_{12} = \Omega_{13} = \cdots = \Omega_{1,T-1} = \Omega_{1T}$$

$$\Omega_{23} = \cdots = \Omega_{2,T-1} = \Omega_{2T}$$

:

$$\Omega_{T-2,T-1} = \Omega_{T-2,T}$$

$$(III.6-2) \quad \Omega_{11}-\Omega_{12} = \Omega_{22}-\Omega_{23} = \cdots = \Omega_{T-1,T-1}-\Omega_{T-1,T}$$

$$(III.6-3) \quad \Omega_{11} = \Omega_{22} = \Omega_{33} = \cdots = \Omega_{T-1,T-1} = \Omega_{TT}$$

All these conditions in (III.6) can be expressed by:

$$(III.7-1) \quad E[y_{it}(u_{i,s+1}-u_{is})] = 0$$

$$0 \leq t \leq T-2, \quad s > t$$

$$(III.7-2) \quad E[y_{it}(u_{i,t+1}-u_{it})-y_{i,t+1}(u_{i,t+2}-u_{i,t+1})] = 0$$

$$1 \leq t \leq T-2$$

$$(III.7-3) \quad E[\bar{u}_i(u_{i,t+1}-u_{it})] = 0$$

$$1 \leq t \leq T-1$$

where $\bar{u}_i = (1/T)e_T'u_i$. For the derivation of (III.7), see APPENDIX B. (III.7-1) implies $(\frac{1}{2})T(T-1)$ restrictions which are exactly identical to those used in the conventional IV approaches. (III.7-2) shows $(T-2)$ missing linear moment conditions which are not used in the former studies.

(III.7-3) shows $(T-1)$ restrictions that are nonlinear in the sense that δ appears in both \bar{u}_i $(=(1/T)e_T'(y_i - \delta y_{i,-1}))$ and $u_{i,t+1} - u_{it}$ $(=(y_{i,t+1} - y_{it}) - \delta(y_{it} - y_{i,t-1}))$. Again, there are in total $(1/2)T(T-1) + (2T-3)$ restrictions. Table 10 summarizes the total number of moment conditions implied by (SA).

Table 10

| Types of Conditions | (III.7-1) | (III.7-2) | (III.7-3) | Total |
|---------------------|-----------------------|-----------|-----------|--------------------------------|
| T | | | | |
| 2 | 1 | 0 | 1 | 2 |
| 3 | 3 | 1 | 2 | 6 |
| 4 | 6 | 2 | 3 | 11 |
| 5 | 10 | 3 | 4 | 17 |
| T | $(\frac{1}{2})T(T-1)$ | T-2 | T-1 | $(\frac{1}{2})T(T-1) - (2T-3)$ |

Referring to Table 10, we see that the total number of all the available moment conditions is almost twice as big as that of the usual conditions, when T is small. This means that the GMM estimator based on (III.7) could have a substantial efficiency gain over the conventional IV estimators if T is relatively small.

Until now, we have investigated the moment restrictions under the usual assumptions. Here, we consider three cases in which some assumptions are relaxed.

CASE I. Keep all the assumptions except (SA.3)

In this case, the heteroskedasticity of the ϵ 's is

allowed. One might think that the moment restrictions given in (III.7-1) would be all we can have because (III.7-2) and (III.7-3) are based on the homoskedasticity of the ϵ 's. However, this is not the case, because we still have some nonlinear restrictions. To see why, observe the form of Ω in CASE I:

$$\begin{aligned}
 \text{(III.8) } \Omega &= \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \cdots & \Omega_{0T} \\ & \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1T} \\ & & \Omega_{22} & \cdots & \Omega_{2T} \\ & & & \ddots & \\ & & & & \Omega_{TT} \end{bmatrix} \\
 &= \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} & \sigma_{0\alpha} & \cdots & \sigma_{0\alpha} \\ & \sigma_{\alpha}^2 + \sigma_{11} & \sigma_{\alpha}^2 & \cdots & \sigma_{\alpha}^2 \\ & & \sigma_{\alpha}^2 + \sigma_{22} & \cdots & \sigma_{\alpha}^2 \\ & & & \ddots & \vdots \\ & & & & \sigma_{\alpha}^2 + \sigma_{TT} \end{bmatrix}
 \end{aligned}$$

This shows that Type I and III restrictions are still available. The total number of conditions must be $(\frac{1}{2})T(T-1) + (T-2)$. That is, we have an extra $(T-2)$ conditions in comparison with those adopted in the conventional IV approaches. To show the explicit form of the restrictions, let us rearrange them in a slightly different manner:

$$\begin{aligned}
 \text{(III.9-1) } \Omega_{01} &= \Omega_{02} = \Omega_{03} = \cdots = \Omega_{0,T-1} = \Omega_{0T} \\
 &= \Omega_{12} = \Omega_{13} = \cdots = \Omega_{1,T-1} = \Omega_{1T} \\
 &= \Omega_{23} = \cdots = \Omega_{2,T-1} = \Omega_{2T} \\
 &\vdots \\
 &= \Omega_{T-2,T-1} = \Omega_{T-2,T}
 \end{aligned}$$

$$(III.9-2) \quad \Omega_{12} = \Omega_{23} = \Omega_{34} = \cdots = \Omega_{T-1,T}$$

(III.9-1) is just identical to (III.7-1), and therefore, it generates the conventional moment restrictions given in (II.5). (III.9-2) can be expressed by:

$$(III.10) \quad E[u_{12}(u_{13}-u_{11})] = E[u_{13}(u_{14}-u_{12})] \\ = \cdots = E[u_{1,T-1}(u_{1T}-u_{1,T-2})] = 0$$

which are (T-2) nonlinear restrictions. If the ϵ 's are heteroskedastic, we do not have any extra linear restrictions. However, the existence of the nonlinear conditions given in (III.10) again prevents the conventional IV estimators from being efficient.

CASE II. Keep (SA.2) and (SA.3)

Here, we allow the ϵ 's to be correlated with each other. For example, the ϵ 's could follow an AR or/and MA process. In this situation, it would be unreasonable to assume that y_{i0} is independent of the ϵ 's. The conventional IV estimators can not be consistent, because they use invalid instrument variables, in the sense that $E[y_{it}(u_{i,s+1}-u_{is})] \neq 0$ for $s > t$. However, we can find some available nonlinear moment conditions. To see this, observe the form of Ω under CASE II:

$$(III.10) \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \cdots & \Omega_{0T} \\ & \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1T} \\ & & \Omega_{22} & \cdots & \Omega_{2T} \\ & & & \ddots & \\ & & & & \Omega_{TT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} + \sigma_{01} & \sigma_{0\alpha} + \sigma_{02} & \cdots & \sigma_{0\alpha} + \sigma_{0T} \\ & \sigma_\alpha^2 + \sigma_\epsilon^2 & \sigma_\alpha^2 + \sigma_{12} & \cdots & \sigma_\alpha^2 + \sigma_{1T} \\ & & \sigma_\alpha^2 + \sigma_\epsilon^2 & \cdots & \sigma_\alpha^2 + \sigma_{2T} \\ & & & \ddots & \\ & & & & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{bmatrix}$$

As we see in (III.10), Type II restrictions are still available. These (T-1) conditions can be written as follows:

$$\begin{aligned} \text{(III.11)} \quad E[(u_{i2} + u_{i1})(u_{i2} - u_{i1})] &= E[(u_{i3} + u_{i2})(u_{i3} - u_{i2})] = \cdots \\ &= E[(u_{i,T-1} + u_{iT})(u_{i,T-1} - u_{iT})] = 0 \end{aligned}$$

These restrictions are all we can have under the intertemporal correlations among the ϵ 's. However, if we assume the stationarity of the ϵ 's, that is, if $E(\epsilon_{is}\epsilon_{i,s+j}) = E(\epsilon_{it}\epsilon_{i,t+j})$ for any t, s , and j , we can have more restrictions. Under the stationarity of the ϵ 's, $\sigma_{s,s+j} = \sigma_{t,t+j}$. Applying this fact to Ω gives us the following extra $(\frac{1}{2})(T-1)(T-2)$ conditions:

$$\begin{aligned} \text{(III.12)} \quad \Omega_{12} &= \Omega_{23} = \Omega_{34} = \cdots = \Omega_{T-2,T-1} = \Omega_{T-1,T} \\ \Omega_{13} &= \Omega_{24} = \cdots = \Omega_{T-3,T-1} = \Omega_{T-2,T} \\ &\vdots \\ \Omega_{1,T-1} &= \Omega_{2T} \end{aligned}$$

Therefore, under the stationarity condition about the ϵ 's, we can have in total $(\frac{1}{2})T(T-1)$ restrictions available for GMM estimation.

CASE III. Keep (SA.3) and (SA.4)

Since y_{i0} could be correlated with ϵ 's, any lagged value of y_{it} can not be used as a legitimate instrument for the differenced equations, because $E[y_{it}(u_{i,s+1} - u_{is})] \neq 0$ for $s > t$ if

neither y_{i0} nor α_i have equicovariances with ϵ_{it} for any t . However, we can have many nonlinear restrictions. For CASE III, Ω takes the following form:

$$(III.13) \quad \Omega = \text{Cov} \begin{bmatrix} y_0 \\ u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \Omega_{02} & \cdots & \Omega_{0T} \\ & \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1T} \\ & & \Omega_{22} & \cdots & \Omega_{2T} \\ & & & \ddots & \\ & & & & \Omega_{TT} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} + \sigma_{01} & \sigma_{0\alpha} + \sigma_{02} & \cdots & \sigma_{0\alpha} + \sigma_{0T} \\ & \sigma_\alpha^2 + \sigma_\epsilon^2 + 2\sigma_{\alpha 1} & \sigma_\alpha^2 + \sigma_{\alpha 1} + \sigma_{\alpha 2} & \cdots & \sigma_\alpha^2 + \sigma_{\alpha 1} + \sigma_{\alpha T} \\ & & \sigma_\alpha^2 + \sigma_\epsilon^2 + 2\sigma_{\alpha 2} & \cdots & \sigma_\alpha^2 + \sigma_{\alpha 1} + \sigma_{\alpha T} \\ & & & \ddots & \vdots \\ & & & & \sigma_\alpha^2 + \sigma_\epsilon^2 + 2\sigma_{\alpha T} \end{bmatrix}$$

Note that $\Omega_{ss} + \Omega_{tt} - 2\Omega_{st} = 2\sigma_\epsilon^2$. Explicitly, this means that $E[(u_{is} - u_{it})^2]$ is the same for any t and s . Therefore, there are $(\frac{1}{2})T(T-1)-1$ restrictions. Effectively, these represent the fact that $\text{Var}(u_{it} - \bar{u}_i) = [(T-1)/T]\sigma_\epsilon^2$ for any t , and $\text{Cov}(u_{it} - \bar{u}_i, u_{is} - \bar{u}_i) = -T\sigma_\epsilon^2$ for $s \neq t$. To clarify this point, consider the transformed error term:

$$(III.14) \quad \begin{bmatrix} u_{i1} - \bar{u}_i \\ u_{i2} - \bar{u}_i \\ \vdots \\ u_{iT} - \bar{u}_i \end{bmatrix} = Qu_i$$

Qu_i has the covariance matrix

$$(III.15) \quad \text{Cov}(Qu_i)$$

$$= \begin{bmatrix} [(T-1)/T]\sigma_{\epsilon}^2 & -\sigma_{\epsilon}^2/T & -\sigma_{\epsilon}^2/T & \cdots & -\sigma_{\epsilon}^2/T \\ & [(T-1)/T]\sigma_{\epsilon}^2 & -\sigma_{\epsilon}^2/T & \cdots & -\sigma_{\epsilon}^2/T \\ & & & \vdots & \\ & & & & [(T-1)/T]\sigma_{\epsilon}^2 & -\sigma_{\epsilon}^2/T \\ & & & & & [(T-1)/T]\sigma_{\epsilon}^2 \end{bmatrix}$$

Note that the rank of $\text{Cov}(\mathbf{Qu}_i)$ is $(T-1)$. Therefore, for the derivation of linearly independent moment conditions only, the last row and column must be ignored. The restrictions implied in (III.15) can be expressed by

$$\begin{aligned} \text{(III.16)} \quad \text{Var}(u_{i1}-\bar{u}_i) &= \text{Var}(u_{i2}-\bar{u}_i) = \cdots = \text{Var}(u_{iT}-\bar{u}_i) \\ &= -(T-1)\text{Cov}(u_{i1}-\bar{u}_i, u_{i2}-\bar{u}_i) = \cdots = -(T-1)\text{Cov}(u_{iT}-\bar{u}_i, u_{iT}-\bar{u}_i) \\ &\quad : \\ &= -(T-1)\text{Cov}(u_{iT-2}-\bar{u}_i, u_{iT-1}-\bar{u}_i) \end{aligned}$$

The number of restrictions in (III.16) is again $(\frac{1}{2})T(T-1)-1$.

IV. Estimation

In Section III, I have introduced a systematic method to derive all the impossible moment conditions from a given set of assumptions. In this section, I examine the GMM estimation procedure based on those restrictions. Even though using all the moment conditions will no doubt improve the efficiency of the GMM estimator, the existence of the nonlinear restrictions makes a simple IV treatment impossible. One of the greatest advantages of the conventional IV methods is that the estimators could be obtained using standard IV (2SLS) software. Therefore, one may be reluctant to use nonlinear

GMM methods. However, we can avoid this problem by the linearized GMM procedure; see Newey [1985].

Define the following expressions:

$$B_{1i} = \begin{bmatrix} -y_{i1} & 0 \\ y_{i1}+y_{i2} & -y_{i2} \\ -y_{i2} & y_{i2}+y_{i3} \\ 0 & -y_{i3} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ -y_{i,T-2} \\ y_{i,T-2}+y_{i,T-1} \\ -y_{i,T-1} \end{matrix}$$

$$B_{2i} = \begin{bmatrix} -\bar{u}_i & 0 \\ \bar{u}_i & -\bar{u}_i \\ 0 & \bar{u}_i \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \\ \\ -\bar{u}_i \\ -\bar{u}_i \end{matrix}$$

$$D_{2i} = \begin{bmatrix} (u_{i2}-u_{i1})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \\ (u_{i2}-u_{i3})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \\ \vdots & \vdots & \ddots & \vdots \\ (u_{i2}-u_{i3})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \end{bmatrix}$$

$$H_i = (A_i \ B_{1i} \ B_{2i}); \ D_i = (0 \ 0 \ D_{2i});$$

$$B_1 = \begin{bmatrix} B_{11} \\ B_{12} \\ \vdots \\ B_{1N} \end{bmatrix}; \ B_2 = \begin{bmatrix} B_{21} \\ B_{22} \\ \vdots \\ B_{2N} \end{bmatrix}; \ D_1 = \begin{bmatrix} D_{21} \\ D_{22} \\ \vdots \\ D_{2N} \end{bmatrix};$$

$$H = (A \ B_1 \ B_2); \ D = (0 \ 0 \ D_2); \ H'u = \sum_{i=1}^N H_i'u_i.$$

where B_{1i} is a $T \times (T-2)$ matrix; B_{2i} and D_{2i} are $T \times (T-1)$ matrices. Then, the moment conditions given in (III.7) can be expressed by

$$(IV.1) \quad E(H_i' u_i) = 0, \text{ or}$$

$$E(H'u) = 0$$

A consistent estimator of $C = \text{Cov}(H'u)$ is required to derive the GMM estimator. For this, we can use $\hat{\delta}_A$ to evaluate H_i and u_i . (Let \hat{H}_i and \hat{u}_i be H_i and u_i evaluated at $\hat{\delta}_A$.) Then, $\hat{C} = \frac{1}{N} \sum_{i=1}^N \hat{H}_i' \hat{u}_i \hat{u}_i' \hat{H}_i$ is consistent for C . One may think of $\hat{H}'\hat{H}$ as a consistent estimator of C in the same sense that $A'A$ is consistent for $\text{Cov}(A'u)$. However, $H'H$ is inconsistent unless $T=2$, because $\text{Cov}(A'u)$ includes the fourth-moments of ϵ_{it} . When the ϵ_{it} 's are normally distributed, APPENDIX C shows

$$(IV.2) \quad \text{plim}(1/N) \text{Cov}(H'u) = \sigma_\epsilon^2 \text{plim}(1/N) H'H + \sigma_\epsilon^2 J^*$$

where J^* is a matrix with the form of

$$J^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & J & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and J is a $(T-2) \times (T-2)$ matrix of the form

$$J = \sigma_\epsilon^2 \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \\ & & & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

Therefore, $\hat{H}'\hat{H} + N\hat{J}^*$, not $\hat{H}'\hat{H}$, could be used for C .² In spite of this fact, I believe that \hat{C} is a better estimate for C , because $\hat{H}'\hat{H} + N\hat{J}^*$ is inconsistent under our modified assumptions in (MA).

The GMM estimator, $\bar{\delta}_{\text{GMM}}$, minimizes

$$(IV.2) \quad (u'H)\hat{C}^{-1}(H'u) = [(Y - \delta Y_{-1})' H(\delta)] \hat{C}^{-1} [H(\delta)' (Y - \delta Y_{-1})]$$

where $H(\delta)$ means that H is a function of δ . It has the following asymptotic distribution:

$$(IV.3) \sqrt{N}(\tilde{\delta}_{GMM} - \delta) \rightarrow N(0, \text{plim}[(1/N)(\partial(u'H)/\partial\delta)C^{-1}(\partial(H'u)/\partial\delta)]^{-1})$$

It can be easily shown that

$$(IV.4) \partial(H'u)/\partial\delta = -(H+D)'y_{-1}$$

Therefore, the asymptotic distribution of $\tilde{\delta}_{GMM}$ is given by:

$$(IV.5) \sqrt{N}(\tilde{\delta}_{GMM} - \delta) \rightarrow N(0, \text{plim}[(1/N)y_{-1}'(H+D)C^{-1}(H+D)'y_{-1}]^{-1})$$

and the asymptotic covariance matrix of $\tilde{\delta}_{GMM}$ is evaluated as

$$(IV.6) [y_{-1}'(\tilde{H}+\tilde{D})\hat{C}^{-1}(\tilde{H}+\tilde{D})'y_{-1}]^{-1}$$

where \tilde{H} and \tilde{D} are evaluated at $\tilde{\delta}_{GMM}$.

To examine the characteristics of $\tilde{\delta}_{GMM}$, consider the first order condition for minimization of (IV.2):

$$(IV.7) y_{-1}'(H+D)\hat{C}^{-1}H'(y-\delta y_{-1}) = 0$$

Then, through some algebraic operations, we can show

$$\begin{aligned} (IV.8) \quad \tilde{\delta}_{GMM} &= [y_{-1}'(\tilde{H}+\tilde{D})\hat{C}_{-1}\tilde{H}'y_{-1}]^{-1}y_{-1}'(\tilde{H}+\tilde{D})\hat{C}^{-1}\tilde{H}'y \\ &= [y_{-1}'(\tilde{H}+\tilde{D})\hat{C}^{-1}(\tilde{H}+\tilde{D})'y_{-1}]^{-1}y_{-1}'(\tilde{H}+\tilde{D})\hat{C}^{-1}(\tilde{H}+\tilde{D})'(y-P_v\tilde{u}) \end{aligned}$$

where \tilde{u} is evaluated at $\tilde{\delta}_{GMM}$. (IV.8) has an interesting implication. It suggests that an iterative procedure with any initial consistent estimator of δ may generate the GMM estimator. This turns out to be true. Furthermore, only one iteration is needed to get an estimator which has the same asymptotic distribution as $\tilde{\delta}_{GMM}$. To see this, consider a new estimator, $\bar{\delta}$, which replaces \tilde{H} , \tilde{D} , and \tilde{u} in (IV.8) by \hat{H} , \hat{D} , \hat{u} (evaluated, say, at $\hat{\delta}_A$). Then,

$$\begin{aligned} (IV.9) \quad \bar{\delta} &= [y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'y_{-1}]^{-1}y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'(y-P_v\hat{u}) \\ &= \delta + [y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'y_{-1}]^{-1}y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'u \\ &\quad - [y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'y_{-1}]^{-1}y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}\hat{D}'\hat{u} \end{aligned}$$

$$\begin{aligned}
&= \delta + [Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'Y_{-1}]^{-1}Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}H'u \\
&\quad + [Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'Y_{-1}]^{-1}Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}-H)'u \\
&\quad - [Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'Y_{-1}]^{-1}Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}\hat{D}'(\hat{u}-u)
\end{aligned}$$

because $P_v\hat{H}=0$ and $P_v\hat{D}=\hat{D}$. Observe that

$$\hat{u}_1 - \bar{u}_1 = -(\hat{\delta}_A - \delta)(1/T)e_T'y_{i,-1}.$$

Therefore,

$$\begin{aligned}
&\text{(IV.10)} \\
\hat{B}_{2i} - B_{2i} &= -(\hat{\delta}_A - \delta) \begin{bmatrix} -(1/T)e_T'y_{i,-1} & 0 \\ (1/T)e_T'y_{i,-1} & -(1/T)e_T'y_{i,-1} \\ 0 & (1/T)e_T'y_{i,-1} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \vdots \\ -(1/T)e_T'y_{i,-1} \\ (1/T)e_T'y_{i,-1} \end{bmatrix}
\end{aligned}$$

Note that

$$\begin{aligned}
&\text{(IV.11)} \begin{bmatrix} -(1/T)e_T'y_{i,-1} & 0 \\ (1/T)e_T'y_{i,-1} & -(1/T)e_T'y_{i,-1} \\ 0 & (1/T)e_T'y_{i,-1} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{bmatrix}' \begin{bmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \\ \vdots \\ u_{iT} \end{bmatrix} \\
&= \begin{bmatrix} (u_{i2}-u_{i1})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \\ (u_{i2}-u_{i1})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \\ \vdots & \vdots & \vdots & \vdots \\ (u_{i2}-u_{i1})/T & (u_{i3}-u_{i2})/T & \cdots & (u_{iT}-u_{i,T-1})/T \end{bmatrix}' \begin{bmatrix} y_{i0} \\ y_{i1} \\ \vdots \\ y_{i,T-1} \end{bmatrix} \\
&= D_{2i}'y_{i,-1}
\end{aligned}$$

Using (IV.10) and (IV.11), we have

$$\begin{aligned}
&\text{(IV.12)} \quad (\hat{B}_{2i} - B_{2i})'u = -(\hat{\delta}_A - \delta)D_{2i}'y_{-1} \quad \text{and} \\
&\quad (\hat{H} - H)'u = -(\hat{\delta}_A - \delta)D'y_{-1}
\end{aligned}$$

Substituting (IV.12) into (IV.9) gives

$$\text{(IV.13)}$$

$$\bar{\delta} - \delta = [Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}(\hat{H}+\hat{D})'Y_{-1}]^{-1}Y_{-1}'(\hat{H}+\hat{D})\hat{C}^{-1}\hat{H}'u$$

+ $(\hat{\delta}_A - \delta) [Y_{-1}' (\hat{H} + \hat{D}) \hat{C}^{-1} (\hat{H} + \hat{D})' Y_{-1}]^{-1} Y_{-1}' (\hat{H} + \hat{D}) \hat{C}^{-1} (\hat{D} - D)' Y_{-1}$
 because $\hat{D}'(\hat{u} - u) = -(\hat{\delta}_A - \delta) \hat{D}' Y_{-1}$. The second term in the right
 side of (IV.13) is asymptotically negligible;

$$\text{plim}(1/\sqrt{N}) (\hat{D} - D)' Y_{-1} (\hat{\delta}_A - \delta) = \text{plim} \sqrt{N} (\hat{\delta}_A - \delta) \text{plim}(1/N) (\hat{D} - D)' Y_{-1} = 0$$

Note that

$$\begin{aligned} & [Y_{-1}' (\hat{H} + \hat{D}) \hat{C}^{-1} (\hat{H} + \hat{D})' Y_{-1}]^{-1} Y_{-1}' (\hat{H} + \hat{D}) \hat{C}^{-1} H' u / \sqrt{N} \\ & \rightarrow N(0, \text{plim}[(1/N) Y_{-1}' (H + D) C^{-1} (H + D)' Y_{-1}]^{-1}) \end{aligned}$$

This means that $\bar{\delta}$ is asymptotically identical to $\tilde{\delta}_{\text{GMM}}$.

This result is actually not surprising. Newey [1985] develops a simple linearized GMM estimator which has the same asymptotic distribution as the nonlinear GMM estimator. His method is applicable with any initial consistent estimator. Applying Newey's formula [1985, p. 238] to our model generates
 (IV.14) $\hat{\delta}_A + [Y_{-1}' (\hat{H} + \hat{D}) C^{-1} (\hat{H} + \hat{D})' Y_{-1}]^{-1} Y_{-1}' (\hat{H} + \hat{D}) C^{-1} \hat{H}' \hat{u}$
 which can be shown algebraically identical to $\bar{\delta}$. Therefore, it turns out that $\bar{\delta}$ is nothing but Newey's linearized GMM.

Even though the efficiency of $\tilde{\delta}_{\text{GMM}}$ is not questionable, one may argue that unless $\tilde{\delta}_{\text{GMM}}$ has a significant efficiency gain, $\hat{\delta}_A$ would be preferred because of its simplicity.

Therefore, it would be worth demonstrating explicitly the efficiency comparison between $\tilde{\delta}_{\text{GMM}}$ and $\hat{\delta}_A$, and seeing in what cases the efficiency gain of $\tilde{\delta}_{\text{GMM}}$ over $\hat{\delta}_A$ is greatest. For simplicity, consider the case in which $T=2$. There are only two moment conditions available:

$$(IV.15) \quad E \begin{bmatrix} Y_{i0} (u_{i2} - u_{i1}) \\ \bar{u}_i (u_{i2} - u_{i1}) \end{bmatrix} = 0$$

where $\bar{u}_i = (\frac{1}{2})(u_{i2} + u_{i1})$. Define

$$A_i = \begin{bmatrix} -y_{i0} \\ -y_{i0} \end{bmatrix} ; B_{2i} = \begin{bmatrix} -\bar{u}_i \\ -\bar{u}_i \end{bmatrix} ; D_{2i} = \begin{bmatrix} (\frac{1}{2})(u_{i2}-u_{i1}) \\ (\frac{1}{2})(u_{i2}-u_{i1}) \end{bmatrix}$$

$$H_i = (A_i \ B_{2i}) ; D_i = (0, \ D_{2i})$$

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{bmatrix} ; B = \begin{bmatrix} B_{11} \\ B_{12} \\ \vdots \\ B_{1N} \end{bmatrix} ; D = \begin{bmatrix} D_{21} \\ D_{22} \\ \vdots \\ D_{2N} \end{bmatrix} ; H = (A, \ B_2)$$

If the ϵ 's are normally distributed with variance σ_ϵ^2 , $\text{Cov}(H'u) = \sigma_\epsilon^2 E(H'H)$. This is true because J^* is irrelevant in (IV.2) when $T=2$. Then,

$$(IV.16) \ \sqrt{N}(\hat{\delta}_{GMM} - \delta) \rightarrow N(0, \sigma_\epsilon^2 \psi_1^{-1})$$

where $\psi_1 = \text{plim}[(1/N)Y_{-1}'(H+D)(H'H)^{-1}(H+D)'Y_{-1}]$. APPENDIX D shows

$$\psi_1 = \text{plim}(1/N)Y_{-1}'P_H Y_{-1} + a$$

where $a = (1/8)\sigma_\epsilon^4/[\sigma_\alpha^2(1-\rho_{0\alpha}^2)+\sigma_\epsilon^2/2] - (1/2)\sigma_\epsilon^2 < 0$. $\rho_{0\alpha}$ denotes the correlation coefficient of y_0 and α , i.e., $\rho_{0\alpha} = \sigma_{0\alpha}/(\sigma_0\sigma_\alpha)$. On the other hand, the conventional IV estimator, $\hat{\delta}_A$, has the

following asymptotic distribution:

$$(IV.17) \ \sqrt{N}(\hat{\delta}_A - \delta) \rightarrow N(0, \sigma_\epsilon^2 \psi_2^{-1})$$

where

$$\psi_2 = \text{plim}(1/N)Y_{-1}'P_A Y_{-1}$$

$$= \text{plim}(1/N)Y_{-1}'P_H Y_{-1} + \text{plim}(1/N)Y_{-1}'M_A B_2 (B_2'M_A B_2)^{-1} B_2'M_A Y_{-1}.$$

APPENDIX D also shows

$$\text{plim}(1/N)Y_{-1}'M_A B_2 (B_2'M_A B_2)^{-1} B_2'M_A Y_{-1} = -(1/2)[\sigma_\alpha^2(1-\rho_{0\alpha}^2)+\sigma_\epsilon^2/2]$$

Then,

$$(IV.18) \ \psi_1 - \psi_2 = (1/2)\sigma_\alpha^4(1-\rho_{0\alpha}^2)^2/[\sigma_\alpha^2(1-\rho_{0\alpha}^2)+\sigma_\epsilon^2/2] > 0$$

By observing (II.18), we can see that the size of efficiency gain of $\tilde{\delta}_{GMM}$ over $\hat{\delta}_A$ depends on σ_ϵ^2 , σ_α^2 , and $\rho_{0\alpha}$. The greater σ_α^2 is and the smaller $\rho_{0\alpha}^2$ and σ_ϵ^2 are, the greater efficiency gain $\tilde{\delta}_{GMM}$ has.

Another possible problem in our estimation method is that $\tilde{\delta}_{GMM}$ loses its consistency under the heteroskedasticity of ϵ 's while $\hat{\delta}_A$ does not. Therefore, an appropriate test procedure may be required. Since $\tilde{\delta}_{GMM}$ is efficient under the null hypothesis (homoskedasticity of ϵ 's), and $\hat{\delta}_A$ is consistent under both null and alternative hypotheses, Hausman's [1978] misspecification test is available. Under the null hypothesis, the test statistic,

$$(II.19) \quad (\tilde{\delta}_{GMM} - \hat{\delta}_A)' (M_2 - M_1)^{-1} (\tilde{\delta}_{GMM} - \hat{\delta}_A)$$

converges in distribution to $\chi^2(T-1)$, where M_1 and M_2 are the asymptotic variances of $\tilde{\delta}_{GMM}$ and $\hat{\delta}_A$, respectively. Note that $\hat{\delta}_A$ is not efficient even when the null hypothesis is rejected. This is because $A'A$ is no longer consistent for $Cov(A'u)$ and because, as CASE I demonstrates in Section III, we still have extra $(T-2)$ moment conditions; see (III.9). To obtain an efficient estimator, we would need to construct the GMM estimator based on the moment conditions in (III.9). Again, Newey's method could be used to generate the linearized GMM estimator.

V. Estimation with Exogenous Variables

Until now, we have investigated the GMM estimation method for the simple dynamic model with a one-period lagged

dependent variable as the only explanatory variable. In this section, we extend the results obtained in the previous sections to the general model which includes exogenous variables. The most surprising finding is that there exist an enormously large number of moment conditions available for the dynamic model. Many former studies for the static model assume strong exogeneity conditions, in the sense that all the exogenous variables are uncorrelated with the ϵ 's at all leads and lags. These conditions, however, do not generate extra instruments for the static model. On the contrary, all these strong exogeneity conditions must be considered in the dynamic model. I will explain this first, and later include the moment conditions obtained in the previous sections.

The model can be written as

$$\begin{aligned} \text{(V.1-1)} \quad y_{it} &= \delta y_{i,t-1} + x_{it}\beta + z_i\gamma + (\alpha_i + \epsilon_{it}) \\ &= \delta y_{i,t-1} + x_{it}\beta + z_i\gamma + u_{it} \\ &\quad i=1,2,\dots,N; t=1,2,\dots,T \end{aligned}$$

where x_{it} is a $1 \times k$ row vector of time-varying exogenous variables, and z_i is a $1 \times g$ row vector of time-invariant exogenous variables. That is, x_{it} and z_i are stochastically independent of ϵ_{it} . For the i th cross-section, (V.1-1) can be rewritten as

$$\text{(V.1-2)} \quad y_i = \delta y_{i,-1} + X_i\beta + Z_i\gamma + u_i$$

For all the observations, the model can be expressed in matrix form as:

$$\begin{aligned} \text{(V.1-3)} \quad y &= \delta y_{-1} + X\beta + Z\gamma + u \\ &= (y_{-1}, X, Z)(\delta, \beta', \gamma')' + u \end{aligned}$$

$$= W\theta + u$$

For the static model (no y_{-1} present), Hausman and Taylor [1981], Amemiya and MacCurdy [1986], and Breusch, Mizon, and Schmidt [1989] develop simple IV estimation methods, each of which provides a consistent and efficient estimator under a given set of exogeneity conditions between x , z and α . To see those exogeneity restrictions, it would be useful to partition X and Z as follows:

$$(V.2-1) \quad X = (X_1, X_2)$$

$$(V.2-2) \quad Z = (Z_1, Z_2)$$

where X_1 and Z_1 are uncorrelated with α while X_2 and Z_2 are correlated with α . For notational convenience, define a $NT \times NTm$ matrix, S^* , as follows:

$$(V.3) \quad S = \begin{bmatrix} S_{11} \\ S_{12} \\ \vdots \\ S_{1T} \\ \hline \vdots \\ \vdots \\ \hline S_{N1} \\ S_{N2} \\ \vdots \\ S_{NT} \end{bmatrix} ; \quad S^* = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1T} \\ S_{11} & S_{12} & \cdots & S_{1T} \\ \vdots & \vdots & & \vdots \\ S_{11} & S_{12} & \cdots & S_{1T} \\ \hline \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \hline S_{N1} & S_{N2} & \cdots & S_{NT} \\ S_{N1} & S_{N2} & \cdots & S_{NT} \\ \vdots & \vdots & & \vdots \\ S_{N1} & S_{N2} & & S_{NT} \end{bmatrix}$$

where m is the number of columns of S . Note that S^* is time-invariant matrix, in the sense that each of its columns is time-invariant. If we follow the interpretation provided by Breusch, Mizon, and Schmidt, the exogeneity conditions imposed by these authors can be compactly described by

$$(V.4) \quad \text{plim}(1/N)G'u = 0$$

where $G=(Q_vX, P_vR)$. R is defined as follows:

$$(V.5) \quad R = \begin{cases} (X_1, Z_1) & \text{for Hausman and Taylor} \\ (X_1, Z_1, X_1^*) & \text{for Amemiya and MaCurdy} \\ (X_1, Z_1, X_1^*, (Q_vX_2)^*) & \text{for Breusch, Mizon, and Schmidt} \end{cases}$$

Since the comparison of these three approaches is not the task of this paper, we will use R without preference to any particular approach. A simple two stage least square (2SLS) treatment using G as instrument variables will provide a consistent estimator of θ . Note that $E(uu') = \sigma_\epsilon^2 \Gamma = \sigma_\epsilon^2 (Q_v + \phi^2 P_v)$ where $\phi^2 = (\sigma_\epsilon^2 + T\sigma_\alpha^2) / \sigma_\epsilon^2$. The simple 2SLS estimator, $\hat{\theta}_G$ is the IV estimator applied to the following transformed equation:

$$(V.6) \quad \Gamma^{1/2}y = \Gamma^{1/2}W\theta + \Gamma^{1/2}u$$

Then, we have

$$\hat{\theta}_G = (W'\Gamma^{-1/2}G'(G'G)^{-1}G'\Gamma^{-1/2}W)^{-1}W'\Gamma^{-1/2}G(G'G)^{-1}G'\Gamma^{-1/2}y$$

Through some algebraic operation, it can also be shown that³

$$(V.7) \quad \hat{\theta}_G = (W'G(G'RG)^{-1}G'W)^{-1}W'G(G'RG)^{-1}G'y$$

The asymptotic distribution of $\hat{\theta}_G$ is given:

$$(V.8) \quad \sqrt{N}(\hat{\theta}_G - \theta) \rightarrow N(0, \sigma_\epsilon^2 \text{plim}[(1/N)W'G(G'RG)^{-1}G'W]^{-1}).$$

$\hat{\theta}_G$ is also a GMM estimator. To see why, note that $\text{Cov}(G'u) = \sigma_\epsilon^2 G'RG$. Then, the GMM estimator based on the exogeneity conditions in (V.4) minimizes

$$(y - W\theta)'G(G'RG)^{-1}G'(y - W\theta)$$

and it is exactly identical to $\hat{\theta}_G$.

Even though (V.4) could incorporate all the exogeneity conditions on x and z in relation with α , it does not exploit

$$(V.9-2) \quad E(z_i' \epsilon_{iR}) = 0$$

where $t, s = 1, 2, \dots, T$. In terms of u_{is} , (V.9) implies:

$$(V.10-1) \quad E(x_{it}'(u_{i,s+1}-u_{is})) = 0$$

$$(V.10-2) \quad E(z_i' (u_{i,s+1} - u_{i,s})) = 0$$

where $s=1,2,\cdots,T-1$. Define

$$X_i^{**} = \begin{bmatrix} -x_{i1} & 0 & & & & & \\ x_{i1} & -x_{i1} & & & & & \\ 0 & x_{i1} & & & & & \\ 0 & 0 & & & & & \\ 0 & 0 & & & & & \\ \vdots & \vdots & & & & & \\ 0 & 0 & & -x_{i1} & & & \\ 0 & 0 & & x_{i1} & & & \end{bmatrix} \quad \dots \quad \begin{bmatrix} -x_{iT} & 0 & & & & & \\ x_{iT} & -x_{iT} & & & & & \\ 0 & x_{iT} & & & & & \\ 0 & 0 & & & & & \\ 0 & 0 & & & & & \\ \vdots & \vdots & & & & & \\ 0 & 0 & & & & & -x_{iT} \\ 0 & 0 & & & & & x_{iT} \end{bmatrix}$$

where X_i^{**} is a $T \times (kT(T-1))$ matrix and Z_i^{**} is a $T \times (g(T-1))$ matrix. Let

$$X^{**} = \begin{bmatrix} x_1^{**} \\ \vdots \\ x_N^{**} \end{bmatrix}; \quad Z^{**} = \begin{bmatrix} z_1^{**} \\ \vdots \\ z_N^{**} \end{bmatrix}; \quad W^{**} = (X^{**}, Z^{**})$$

Then, we may use W^{**} as instruments, because $\text{plim}(1/N)W^{**}'u=0$.

Define $F=(W^{**} \ G)=(X^{**} \ Z^{**} \ Q_v X \ P_v R)$. Then, all the available exogeneity conditions can be compactly described by

$$(V.11) \text{plim}(1/N)F'u = E(F'u) = 0$$

Note that not all the moment restrictions in (V.11) are linearly independent. To see why, we can rewrite the restrictions in (V.11) separately as follows:

$$(V.12-1) E(X^{**}u) = 0$$

$$(V.12-2) E(Z^{**}u) = 0$$

$$(V.12-3) E(X'Q_v u) = 0$$

$$(V.12-4) E(R'P_v u) = 0$$

(V.9-1) implies $E[x_{it}'(\epsilon_{it} - \bar{\epsilon}_i)] = E[x_{it}'(u_{it} - \bar{u}_i)] = 0$, which in turn shows (V.12-3). Since (V.12-1) is derived from (V.9-1), (V.12-1) also implies (V.12-3). That is, k restrictions among those in (V.12-1) are already incorporated by (V.12-3). Therefore, (V.12-1) contains extra $[(T^2 - T - 1)k]$, not $[T(T-1)k]$, conditions which are not captured by (V.12-3). Since (V.12-2) implies $[(T-1)g]$ restrictions, (V.11) includes in total extra $[(T^2 - T - 1)k + (T-1)g]$ conditions compared with (V.4).

Following the same reasoning, we can easily show that the rank of $(X^{**} Q_v X)$ is $T(T-1)k$, while the number of columns of $(X^{**} Q_v X)$ is $(T^2 - T + 1)k$. That is, F does not have full-column rank. Therefore, we have to exclude $Q_v X$, or, as I assume here, the last k columns of X^{**} from F , when we proceed the 2SLS estimation based on (V.11). Then, the 2SLS estimator, $\hat{\theta}_F$, has the same form as $\hat{\theta}_G$ except that F replaces G :

$$\begin{aligned} (V.13) \quad \hat{\theta}_F &= (W'\Gamma^{-\frac{1}{2}}F(F'F)^{-1}F'\Gamma^{-\frac{1}{2}}W)^{-1}W'\Gamma^{-\frac{1}{2}}F(F'F)^{-1}F'\Gamma^{-\frac{1}{2}}y \\ &= (W'F(F'\Gamma F)^{-1}F'W)^{-1}W'F(F'\Gamma F)^{-1}F'y \end{aligned}$$

The asymptotic distribution of $\hat{\theta}_F$ is given:

$$(V.14) \sqrt{N}(\hat{\theta}_F - \theta) \rightarrow N(0, \sigma_\epsilon^2 \text{plim}[(1/N)W'F(F'FF)^{-1}F'W]^{-1})$$

$\hat{\theta}_F$ is also the GMM estimator based on all the available exogeneity conditions give in (V.11).

However, W^{**} is irrelevant for the static model. To see why, note that Hausman and Taylor, and Amemiya and MacCurdy, use (Q_v, R) as instruments, not $G=(Q_vX, P_vR)$. It actually does not matter whether we choose (Q_vX, P_vR) or (Q_v, R) , because both generate the same estimator; see Breusch, Mizon and Schmidt [1989]. By the same reasoning, the IV treatment with $F=(W^{**}, Q_vX, P_vR)$ on the static model amounts to that with (W^{**}, Q_v, R) . Now, since the columns of W^{**} are linear combinations of columns of Q_v , in the sense that $W^{**}=Q_vW^{**}$, the projection of a variable onto (W^{**}, Q_v, R) must be equal to that onto (Q_v, R) . This means that (W^{**}, Q_v, R) provides the same estimator as (Q_v, R) , and therefore (Q_vX, P_vR) . In other words, the inclusion of W^{**} into the set of instrument variables is irrelevant for the estimation of the static model.

However, the situation is different for the dynamic model, because Q_v is no longer a legitimate instrument. (This is why the within estimator is inconsistent.) The existence of the lagged dependent variable as an explanatory variable makes (Q_v, R) substantially different from $G=(Q_vX, P_vR)$. These instrument sets no longer generate the same estimator. This is because Q_vX , not Q_v alone, is a legitimate instrument. This fact has a powerful implication. Since W^{**} and Q_vX are linearly independent, the inclusion of W^{**} into the set of instrument variables will improve the efficiency of the 2SLS

estimator for the dynamic model; i.e., $\hat{\theta}_F$ is more efficient than $\hat{\theta}_G$ when y_{-1} is present. Furthermore, when the dimension of W^{**} $((T^2-T-1)k+(T-1)g)$ is considered, the efficiency gain from using W^{**} may be enormously huge. This is an extraordinary case for the general regression model, because the number of instrumental variables based on the exogeneity conditions usually does not change whether the given model is dynamic or static. Therefore, this suggests that the moment conditions under a model using panel data should be carefully counted.

Now, let us include the moment conditions used in the conventional IV approaches for the dynamic model. For this, define

$$S = (A, F)$$

Then, the orthogonality conditions can be expressed by

$$(V.15) \quad \text{plim}(1/N)S'u = 0$$

Since $\text{Cov}(S'u) = \sigma_e^2 E(S'FS)$, $S'FS$ can be used as a consistent estimator of $\text{Cov}(S'u)$. The GMM estimator based on (V.15) is given by:

$$(V.16) \quad \hat{\theta}_S = (W'S(S'FS)^{-1}S'W)^{-1}W'S(S'FS)^{-1}S'y$$

where

$$(V.17) \quad \sqrt{N}(\hat{\theta}_S - \theta) \rightarrow N(0, \sigma_e^2 \text{plim}[(1/N)W'S(S'FG)^{-1}S'W]^{-1})$$

This is also algebraically identical to the IV estimator using S as the instruments on the transformed equation given in (V.6). APPENDIX E shows

$$(V.18) \quad W'S(S'FG)^{-1}S'W = W'F(F'FF)^{-1}F'W + K_1$$

where K_1 is a positive semidefinite matrix. This means that

$\hat{\theta}_S$ is more efficient than $\hat{\theta}_F$.

As a final step, let us include the whole set of moment conditions available. Define

$$L(\theta) = (H(\theta), F)$$

The notation, $H(\theta)$, indicates the fact that, unlike S , H is a function of θ . Then,

$$(V.19) \text{plim}(1/N)H(\theta)'u = 0$$

Note that $\zeta = \text{Cov}(L'u) = \sum_{i=1}^N E(L_i'u_i u_i' L_i)$ where $L_i = (H_i, F_i)$. Evaluate L_i and u_i using any consistent estimator, $\hat{\theta}$, and denote them by \hat{L}_i and \hat{u}_i . Then, $\zeta = \sum_{i=1}^N \hat{L}_i' \hat{u}_i \hat{u}_i' \hat{L}_i$ is a consistent estimator of ζ . The GMM estimator, $\tilde{\theta}_L$, minimizes $(y - W\theta)' L(\theta) \zeta^{-1} L(\theta)' (y - W\theta)$

and its asymptotic distribution is given:

$$(V.20) \sqrt{N}(\tilde{\theta}_L - \theta) \rightarrow N(0, \text{plim}[(1/N)W'(L + L_D)\zeta^{-1}(L + L_D)'W]^{-1})$$

where $L_D = (D, 0, 0)$.

For an efficiency comparison, assume that the ϵ 's are normally distributed. Then, using LEMMA 1 in APPENDIX A, we easily show

$$\begin{aligned} (V.21) \text{plim}(1/N)\text{Cov}(L'u) &= \sigma_\epsilon^2 \text{plim}(1/N) \begin{bmatrix} H'H & H'E & 0 \\ E'H & E'E & 0 \\ 0 & 0 & \phi^2 R'P_V R \end{bmatrix} \\ &\quad + \sigma_\epsilon^2 \begin{bmatrix} J^* & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \sigma_\epsilon^2 \text{plim}(1/N) L' \Gamma L + \sigma_\epsilon^2 J^* \end{aligned}$$

where $E = (W^{**}, Q_V X)$. This implies that $\hat{L}' \Gamma \hat{L} + N \hat{J}^*$ can be used as a consistent estimator of ζ . If we replace $\hat{\zeta}$ by $\hat{L}' \Gamma \hat{L} + N \hat{J}^*$, as shown in APPENDIX F, the inverse of the asymptotic covariance

matrix of $\tilde{\theta}_L$ is given by:

$$(V.22) \quad W'S(S'RS)^{-1}S'W + K_2$$

where K_2 is a positive semidefinite matrix. Therefore, $\tilde{\theta}_L$ is more efficient than $\hat{\theta}_S$. Actually, $\tilde{\theta}_L$ can be said to be efficient, because it exploits all the available moment conditions. The following linearized GMM estimator can also be used:

$$(V.21) \quad \bar{\theta}_L = (W'(\hat{L} + \hat{L}_D)\hat{\zeta}^{-1}(\hat{L} + \hat{L}_D)'W)^{-1}W'(\hat{L} + \hat{L}_D)\hat{\zeta}^{-1}(\hat{L} + \hat{L}_D)'(y - P_v\hat{u})$$

where $\hat{\cdot}$ means "evaluated at a consistent estimator of θ ." Again, $\bar{\theta}_L$ has the same asymptotic distribution as $\tilde{\theta}_L$.

VI. Conclusion

In this paper, I have adopted standard assumptions for the dynamic panel data model, and I have characterized all of the moment conditions that these assumptions imply. I showed that previous IV estimators do not impose all of the available linear moment conditions, and also that there are some nonlinear moment conditions that they do not incorporate. This reveals the inefficiency of the conventional IV estimators. I propose an efficient GMM estimator. Since the GMM estimator is nonlinear, I also considered an efficient linearized version of the GMM estimator.

I also extended this approach to the dynamic model that includes exogenous variables. I showed that the existing treatments of the static model incorporate some but not all of the restrictions that are implied by strong exogeneity of the

exogenous variables. The extra restrictions implied by the strong exogeneity are irrelevant in the static model, but they are relevant in the dynamic model. This fact generates a large number of additional instruments, all of which can be exploited either by a linear IV estimator or by a nonlinear GMM estimator. These extra moment conditions may result in a large gain in efficiency in the dynamic model.

ENDNOTES

1. This assumption implies that y_{i0} 's are stochastic. Since the α 's are random effects, and since y_{i0} may include α_i , this seems reasonable. Also, the assumption, $E(y_{i0})=0$, is adopted for simplicity. Actually, it is enough to assume that Σ is the second-moment matrix:

$$\Sigma = E[(Y_0, \alpha, \epsilon_1, \epsilon_2, \dots, \epsilon_T)' (Y_0, \alpha, \epsilon_1, \epsilon_2, \dots, \epsilon_T)]$$

2. J contains σ_ϵ^2 . Therefore, J^* must be estimated by any consistent estimator of σ_ϵ^2 .
3. We assume here that ϕ^2 is known. For the estimation of ϕ^2 , see Hausman and Taylor [1981].

APPENDICES

APPENDIX A

To derive the GMM estimator based on the restrictions in (II.5), we need to know the covariance matrix of $A_i'u_i$ under (SA) given in Section III. For this, the following lemmas are useful.

LEMMA 1

Suppose that ξ_1 , ξ_2 , and ξ_3 are random variables. Assume that ξ_1 is independent of ξ_3 , and $E(\xi_1)=0$. Also assume that

$$\xi_2 = a\eta_1 + b\eta_2,$$

where η_1 is independent of ξ_3 ; and η_2 is independent of ξ_1 , and a and b are some constants. Then,

$$E(\xi_1\xi_2\xi_3) = E(\xi_1\xi_2)E(\xi_3).$$

Proof.

$$\begin{aligned} E(\xi_1\xi_2\xi_3) &= aE(\xi_1\eta_1\xi_3) + bE(\xi_1\eta_2\xi_3) \\ &= aE(\xi_1\eta_1)E(\xi_3) + bE(\xi_1)E(\eta_2\xi_3) \\ &= aE(\xi_1\eta_1)E(\xi_3) \\ &= E[(a\eta_1 + b\eta_2)\xi_1]E(\xi_3) \\ &= E(\xi_1\xi_2)E(\xi_3) \quad QED \end{aligned}$$

LEMMA 2

$$\begin{aligned} &E[Y_{it}Y_{ih}(\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})] \\ &= E(Y_{it}Y_{ih})E[(\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})], \end{aligned}$$

for $t \leq h$, $t < s, k$.

Proof.

Since $t < s$, $t < k$, y_{it} is independent of $(\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})$.

Obviously, $E(y_{it})=0$. Furthermore, we can decompose y_{ih} into a part that depends on y_{it} and α_i , which is independent of $(\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})$, and a part that depends on $\epsilon_{i,t+1}, \dots, \epsilon_{ih}$, which is independent of y_{it} . Then LEMMA 1 applies and implies the result. QED

THEOREM 1

$$\text{Cov}(A_i' u_i) = \sigma_\epsilon^2 E(A_i' \Gamma_i A_i) = \sigma_\epsilon^2 E(A_i' A_i)$$

where $\Gamma_i = Q + \phi^2 P$ and $\phi^2 = (\sigma_\epsilon^2 + T\sigma_\alpha^2) / \sigma_\epsilon^2$.

Proof.

A typical element of $A_i' u_i$ is $y_{it}(\epsilon_{i,s+1}-\epsilon_{is})$ where $t = 0, 1, 2, \dots, T-2$, and $s > t$. Then the typical element of $\text{Cov}(A_i' u_i) = E(A_i' u_i u_i' A_i)$ must be $E(y_{it} y_{ih} (\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik}))$ where $t \geq h$, $s > t$, and $k > h$. By LEMMA 2,

$$\begin{aligned} E(y_{it} y_{ih} (\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})) \\ = E(y_{it} y_{ih}) E[(\epsilon_{i,s+1}-\epsilon_{is})(\epsilon_{i,k+1}-\epsilon_{ik})] \end{aligned}$$

Using this fact, we can show

$$E(A_i' u_i u_i' A_i) = E[A_i' E(u_i u_i') A_i] = \sigma_\epsilon^2 E[A_i' \Gamma_i A_i]$$

where $\sigma_\epsilon^2 \Gamma_i = E(u_i u_i')$. It is a well-known fact that $\Gamma_i = \sigma_\epsilon^2 (Q + \phi^2 P)$ where $\phi^2 = (\sigma_\epsilon^2 + T\sigma_\alpha^2) / \sigma_\epsilon^2$; see Hausman and Taylor [1981]. Therefore, we have

$$E(A_i' u_i u_i' A_i) = \sigma_\epsilon^2 E[A_i' A_i]. \quad \text{QED}$$

To derive the GMM estimator, we note that

$$\begin{aligned} \text{plim}(1/N) \text{Cov}(A' u) &= \text{plim}(1/N) \sum_{i=1}^N \text{Cov}(A_i' u_i) \\ &= \text{plim}(1/N) \sum_{i=1}^N E(A_i' u_i u_i' A_i) \end{aligned}$$

$$\begin{aligned}
&= \sigma_{\epsilon}^2 \cdot \text{plim}(1/N) \sum_{i=1}^N E[A_i' A_i] \\
&= \sigma_{\epsilon}^2 \cdot \text{plim}(1/N) E(A'A) \\
&= \sigma_{\epsilon}^2 \cdot \text{plim}(1/N) A'A
\end{aligned}$$

This shows that $(A'A)$ could be used as a consistent estimator of $\text{Cov}(A'u)$. (Actually it must be $\hat{\sigma}_{\epsilon}^2(A'A)$ where $\hat{\sigma}_{\epsilon}^2$ is a consistent estimator of σ_{ϵ}^2 . However, $\hat{\sigma}_{\epsilon}^2$ would cancel and therefore is irrelevant for the derivation of the GMM estimator.) By minimizing

$$(u'A)(A'A)^{-1}(A'u) = \{(Y - \delta Y_{-1})'A\}(A'A)^{-1}\{A'(Y - \delta Y_{-1})\},$$

we obtain the GMM estimator of δ , which is exactly identical to the IV estimator given in (II.6).

APPENDIX B

From (III.6-1), we have (with i suppressed):

$$\begin{aligned}
 \text{(A.B-1)} \quad E(Y_0 u_1) &= E(Y_0 u_2) = E(Y_0 u_3) = \cdots = E(Y_0 u_{T-1}) = E(Y_0 u_T) \\
 E(u_1 u_2) &= E(u_1 u_3) = \cdots = E(u_1 u_{T-1}) = E(u_1 u_T) \\
 E(u_2 u_3) &= \cdots = E(u_2 u_{T-1}) = E(u_2 u_T) \\
 &\vdots \\
 E(u_{T-2} u_{T-1}) &= E(u_{T-2} u_T)
 \end{aligned}$$

This set of equations can be rewritten as follows:

$$\begin{aligned}
 \text{(A.B-2)} \quad E[Y_0(u_2 - u_1)] &= E[Y_0(u_3 - u_2)] = \cdots = E[Y_0(u_T - u_{T-1})] = 0 \\
 E[u_1(u_3 - u_2)] &= \cdots = E[u_1(u_T - u_{T-1})] = 0 \\
 &\vdots \\
 E[u_{T-2}(u_T - u_{T-1})] &= 0
 \end{aligned}$$

Multiplying the equations in the first row of (A.B-2) by δ and adding the equations in the second row gives:

$$\text{(A.B-3)} \quad E(Y_1(u_3 - u_2)) = \cdots = E(Y_1(u_T - u_{T-1})) = 0$$

The same procedure for the other equations generates all the moment conditions given in (III.7-1).

Now, we can rewrite (III.6-2) as follows:

$$\text{(A.B-4)} \quad E[u_1(u_2 - u_1)] = E[u_2(u_3 - u_2)] = \cdots = E[u_{T-1}(u_T - u_{T-1})]$$

These restrictions can be equally expressed by

$$\begin{aligned}
 \text{(A.B-5)} \quad E[u_1(u_2 - u_1) - u_2(u_3 - u_2)] &= E[u_2(u_3 - u_2) - u_3(u_4 - u_3)] = \cdots \\
 &= E[u_{T-2}(u_{T-1} - u_{T-2}) - u_{T-1}(u_T - u_{T-1})] = 0
 \end{aligned}$$

Since $E[Y_0(u_2 - u_1)] = E[Y_1(u_3 - u_2)] = 0$ by (A.B-2) and (A.B-3),

$$\begin{aligned}
 E[u_1(u_2 - u_1) - u_2(u_3 - u_2)] + \delta E[Y_0(u_2 - u_1)] - \delta E[Y_1(u_3 - u_2)] \\
 = E[Y_1(u_2 - u_1) - Y_2(u_3 - u_2)] = 0
 \end{aligned}$$

Similarly, we can show that $E[Y_t(u_{t+1} - u_t) - Y_{t+1}(u_{t+2} - u_{t+1})] = 0$

for any $1 \leq t \leq T-2$.

Finally, (III.6-3) implies:

$$\begin{aligned} \text{(A.B-6)} \quad E[(u_2+u_1)(u_2-u_1)] &= E[(u_3+u_2)(u_3-u_2)] = \cdots \\ &= E[(u_T+u_{T-1})(u_T-u_{T-1})] = 0 \end{aligned}$$

Since $\Omega_{23}=\Omega_{24}=\cdots=\Omega_{2T}$ and $\Omega_{13}=\Omega_{14}=\cdots=\Omega_{1T}$ by (III.6-1) and (III.6-2),

$$E(u_3(u_2-u_1)) = E(u_4(u_2-u_1)) = \cdots = E(u_T(u_2-u_1)) = 0$$

Therefore,

$$E[(u_2+u_1)(u_2-u_1)+u_3(u_2-u_1)+u_4(u_2-u_1)+\cdots+u_T(u_2-u_1)] = 0$$

$$E(\bar{u}(u_2-u_1)) = 0$$

where $\bar{u}=(1/T) \sum_{t=1}^T u_t$. In essentially the same way, we can show that all the restrictions in (III.7-3) are justified.

APPENDIX C

The following lemmas and theorems will help to derive the covariance matrix of $H_1'u_1$ under (SA).

LEMMA 3

All the assumptions in (SA) are satisfied. Then, (with i suppressed)

$$E(Y_t Y_h (\epsilon_{s+1} - \epsilon_s) (\epsilon_{h+1} - \epsilon_h)) = E(Y_t Y_h) E((\epsilon_{s+1} - \epsilon_s) (\epsilon_{h+1} - \epsilon_h))$$

where $t=0,1,\dots,T-2$, $h=1,2,\dots,T-1$, and $t < s$.

Proof.

If $t < h$, LEMMA 2 directly shows the result. Consider the case where $h \leq t$. Since $s > t \geq h$, $\epsilon_{s+1} - \epsilon_s$ is independent of $y_t y_h$, and $E(\epsilon_{s+1} - \epsilon_s) = 0$. If $h+1 = s$, ϵ_{h+1} is independent of $y_t y_h$ and ϵ_h is independent of $\epsilon_{s+1} - \epsilon_s$. If $h+1 < s$, $\epsilon_{h+1} - \epsilon_h$ is independent of $\epsilon_{s+1} - \epsilon_s$. Therefore, $\epsilon_{h+1} - \epsilon_s$ can be decomposed in the same way that LEMMA 1 describes. LEMMA 1 again applies and implies the result. *QED*

LEMMA 3 directly implies the following theorem.

THEOREM 2

$$E(A_1'u_1 u_1'B_{11}) = \sigma_\epsilon^2 E(A_1'\Gamma_1 B_{11}) = \sigma_\epsilon^2 E(A_1'B_{11}).$$

The following lemma is helpful to investigate the form of the covariance matrix of $A_1'u_1 u_1'B_{21}$.

LEMMA 4

Under SA,

$$E[y_t \bar{\epsilon} (\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} - \epsilon_k)] = E(y_t \bar{\epsilon}) E[(\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} - \epsilon_k)]$$

where $s > t$, $s, k = 1, 2, \dots, T-1$, and $\bar{\epsilon} = \sum_{t=1}^T \epsilon_t$.

Proof.

Let $e_{s+1} = \sum_{t=1}^T \epsilon_t - (\epsilon_{s+1} + \epsilon_s)$; then $\bar{\epsilon} = (\epsilon_{s+1} + \epsilon_s)/T + e_{s+1}/T$. Therefore, we have

(A.C-1)

$$\begin{aligned} E(y_t \bar{\epsilon} (\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} - \epsilon_k)) &= E[y_t (\epsilon_{s+1} + \epsilon_s) (\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} - \epsilon_k)]/T \\ &\quad + E[y_t e_{s+1} (\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} - \epsilon_k)]/T \\ &= a + b \end{aligned}$$

Note that y_t is independent of ϵ_{s+1} and ϵ_s . Obviously, $E(y_t) = 0$. If $k > t$, $\epsilon_{k+1} - \epsilon_k$ is independent of y_t . If $k = t$, ϵ_{k+1} is independent of y_t , and ϵ_k is independent of $(\epsilon_{s+1} + \epsilon_s)(\epsilon_{s+1} - \epsilon_s)$. If $k < t$, $\epsilon_{k+1} - \epsilon_k$ is independent of $(\epsilon_{s+1} + \epsilon_s)(\epsilon_{s+1} - \epsilon_s)$. That is, $\epsilon_{k+1} - \epsilon_k$ can be decomposed in the same way that LEMMA 1 assumes. Therefore, we have

$$\begin{aligned} \text{(A.C-2) } a &= E[y_t (\epsilon_{k+1} - \epsilon_k)] E[(\epsilon_{s+1} - \epsilon_s) (\epsilon_{s+1} + \epsilon_s)]/T \\ &= 0 \\ &= E[y_t (\epsilon_{s+1} + \epsilon_s)/T] E[(\epsilon_{s+1} - \epsilon_s) (\epsilon_{k+1} \epsilon_k)] \end{aligned}$$

Furthermore, we can easily see that $\epsilon_{s+1} - \epsilon_s$ is independent of y_t and e_{s+1} , and that $E[(\epsilon_{s+1} - \epsilon_s)] = 0$. If $k > s+1$ or $k < s-1$, $\epsilon_{k+1} - \epsilon_k$ is independent of $\epsilon_{s+1} - \epsilon_s$. If $k = s+1$, ϵ_{k+1} is independent of $(\epsilon_{s+1} - \epsilon_s)$, and ϵ_k is independent of y_t and e_{s+1} . If $k = s$, $\epsilon_{k+1} - \epsilon_k$ is independent of $y_t e_{s+1}$. Finally, if $k = s-1$, ϵ_{k+1} is independent of $y_t e_{s+1}$, and ϵ_k is independent of

$(\epsilon_{s+1}-\epsilon_s)(\epsilon_{s+1}-\epsilon_s)$. Then LEMMA 1 applies and we have

$$(A.C-3) \quad b = E(y_t e_{s+1}/T) E[(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)]$$

Substituting (A.C-2) and (A.C-3) into (A.C-1) gives the result. *QED*

LEMMA 5

$E[y_t \alpha(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)] = E(y_t \alpha) E[(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)]$,
for any s, t, k .

Proof.

α is independent of $(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)$ and $E(\alpha)=0$. Note that
(A.C-4) $y_t = [\delta^t y_0 + \alpha(1+\delta+\dots+\delta^{s-1})] + (\epsilon_t + \delta\epsilon_{t-1} + \dots + \delta^{s-1}\epsilon_1)$.
The first term, $[\delta^t y_0 + \alpha(1+\delta+\dots+\delta^{s-1})]$ is independent of $(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)$. The second term, $(\epsilon_t + \delta\epsilon_{t-1} + \dots + \delta^{s-1}\epsilon_1)$ is independent of α . Therefore, LEMMA 1 applies and implies the result. *QED*

The following theorem shows the form of the covariance matrix of $A_i' u_i$ and $B_{2i}' u_i$.

THEOREM 3

$$E(A_i' u_i u_i' B_{2i}) = \sigma_\epsilon^2 E(A_i' \Gamma_i B_{2i}) = \sigma_\epsilon^2 E(A_i' B_{2i}).$$

Proof.

$A_i' u_i u_i' B_{2i}$ has the typical element of $y_t \bar{u}(u_{s+1}-u_s)(u_{k+1}-u_k)$, where $t < s$, which is identical to $y_t(\alpha + \bar{\epsilon})(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)$. Then, LEMMA 4 and 5 imply

$$(A.C-5) \quad E[y_t \bar{u}(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)] = E[y_t \alpha(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)] \\ + E[y_t \bar{\epsilon}(\epsilon_{s+1}-\epsilon_s)(\epsilon_{k+1}-\epsilon_k)]$$

$$\begin{aligned}
&= E(Y_t \alpha) E[(\epsilon_{s+1} - \epsilon_s)(\epsilon_{k+1} - \epsilon_k)] \\
&\quad + E(Y_t \bar{\epsilon}) E[(\epsilon_{s+1} - \epsilon_s)(\epsilon_{k+1} - \epsilon_k)] \\
&= E(Y_t \bar{u}) E[(u_{s+1} - u_s)(u_{k+1} - u_k)].
\end{aligned}$$

Using (A.C-5), we can easily show the result. *QED*

The following lemmas help us prove Theorem 4 which describe the form of $E(B_{1i}' u_i u_i' B_{1i})$.

LEMMA 6

$$E[Y_t^2 (\epsilon_{t+1} - \epsilon_t)^2] = E(Y_t^2) E[(\epsilon_{t+1} - \epsilon_t)^2] + [E(\epsilon^4) - \sigma_\epsilon^2]$$

Proof.

$$\begin{aligned}
E[Y_t^2 (\epsilon_{t+1} - \epsilon_t)^2] &= E(Y_t^2 \epsilon_{t+1}^2) + E(Y_t^2 \epsilon_t^2) \\
&= E(Y_t^2) E(\epsilon_{t+1}^2) + E[(\delta Y_{t-1} + \alpha + \epsilon_t)^2 \epsilon_t^2] \\
&= E(Y_t^2) E(\epsilon_{t+1}^2) \\
&\quad + E[\{(\delta Y_{t-1} + \alpha)^2 + \epsilon_t^2 + 2(\delta Y_{t-1} + \alpha) \epsilon_t\} \epsilon_t^2] \\
&= E(Y_t^2) E(\epsilon_{t+1}^2) + E[(\delta Y_{t-1} + \alpha)^2 \epsilon_t^2] + E(\epsilon_t^4) \\
&= E(Y_t^2) E(\epsilon_{t+1}^2) + E[(\delta Y_{t-1} + \alpha + \epsilon_t)^2] E(\epsilon_t^2) \\
&\quad - E(\epsilon_t^2) E(\epsilon_t^2) + E(\epsilon_t^4) \\
&= E(Y_t^2) E(\epsilon_{t+1}^2) + E(Y_t^2) E(\epsilon_t^2) + [E(\epsilon_t^4) - \sigma_\epsilon^4] \\
&= E(Y_t^2) E[(\epsilon_{t+1} - \epsilon_t)^2] + [E(\epsilon_t^4) - \sigma_\epsilon^4]. \quad \text{QED}
\end{aligned}$$

LEMMA 7

$$\begin{aligned}
&E[Y_t Y_{t+h} (\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+h+1} - \epsilon_{t+h})] \\
&= E(Y_t Y_{t+h}) E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+h+1} - \epsilon_{t+h})] + \sigma_\epsilon^4, \quad h \geq 1.
\end{aligned}$$

Proof.

$$\begin{aligned}
&(\text{A.C-6}) \quad E[Y_t Y_{t+h} (\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+h+1} - \epsilon_{t+h})] \\
&= E[Y_t Y_{t+h} (\epsilon_{t+1} - \epsilon_t) \epsilon_{t+h+1}] - E[Y_t Y_{t+h} (\epsilon_{t+1} - \epsilon_t) \epsilon_{t+h}]
\end{aligned}$$

$$\begin{aligned}
&= -E[Y_t Y_{t+h} (\epsilon_{t+1} - \epsilon_t) \epsilon_{t+h}] \\
&= E(Y_t Y_{t+h} \epsilon_t \epsilon_{t+h}) - E(Y_t Y_{t+h} \epsilon_{t+1} \epsilon_{t+h})
\end{aligned}$$

Note that

$$\begin{aligned}
\text{(A.C-7)} \quad &E(Y_t Y_{t+h} \epsilon_t \epsilon_{t+h}) \\
&= E(Y_t \epsilon_t) E(Y_{t+h} \epsilon_{t+h}) \\
&= \sigma_\epsilon^4
\end{aligned}$$

$$\begin{aligned}
\text{(A.C-8)} \quad &E(Y_t Y_{t+h} \epsilon_{t+1} \epsilon_{t+h}) \\
&= E[Y_t \{ \delta^h Y_t + \alpha(1 + \delta + \dots + \delta^{h-1}) + (\epsilon_{t+h} + \dots + \delta^{h-1} \epsilon_{t+1}) \} \epsilon_{t+1} \epsilon_{t+h}] \\
&= E[Y_t \{ \delta^h Y_t + \alpha(1 + \delta + \dots + \delta^{h-1}) \} \epsilon_{t+1} \epsilon_{t+h}] \\
&\quad + E[Y_t (\epsilon_{t+h} + \dots + \delta^{h-1} \epsilon_{t+1}) \epsilon_{t+1} \epsilon_{t+h}] \\
&= E[Y_t \{ \delta^h Y_t + \alpha(1 + \delta + \dots + \delta^{h-1}) \}] E(\epsilon_{t+1} \epsilon_{t+h}) \\
&= E[Y_t \{ \delta^h Y_t + \alpha(1 + \delta + \dots + \delta^{h-1}) + (\epsilon_{t+h} + \dots + \delta^{h-1} \epsilon_{t+1}) \}] \\
&\quad \cdot E(\epsilon_{t+1} \epsilon_{t+h}) \\
&= E(Y_t Y_{t+h}) E(\epsilon_{t+1} \epsilon_{t+h}) \\
&= -E(Y_t Y_{t+h}) E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+h+1} - \epsilon_{t+h})]
\end{aligned}$$

Substituting (A.C-7) and (A.C-8) into (A.C-6) gives the result. *QED*

We can combine LEMMA 6 and 7 into the following lemma.

LEMMA 8

$$E[Y_t Y_h (\epsilon_{t+1} - \epsilon_t)(\epsilon_{h+1} - \epsilon_h)] = E(Y_t Y_h) E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{h+1} - \epsilon_h)] + a_{th}$$

where $a_{th} = E(\epsilon^4) - \sigma_\epsilon^4$, if $h=t$; $a_{th} = \sigma_\epsilon^4$, elsewhere.

Now we can evaluate $E(B_{1i}' u_i u_i' B_{1i})$.

THEOREM 4

$$\begin{aligned} E(B_{1i}' u_i u_i' B_{1i}) &= \sigma_\epsilon^2 E(B_{1i}' \Gamma_i B_{1i}) + (d + \sigma_\epsilon^4) J_1 \\ &= \sigma_\epsilon^2 E(B_{1i}' B_{1i}) + (d + \sigma_\epsilon^2) J_1 \end{aligned}$$

where

$$J_1 = \begin{bmatrix} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & & & \\ 0 & -1 & 2 & \cdot & & \\ 0 & 0 & -1 & \cdot & & \\ \vdots & \vdots & \vdots & & 2 & -1 \\ 0 & 0 & 0 & & -1 & 2 \end{bmatrix}_{(T-2) \times (T-2)},$$

$$d = E(\epsilon^4) - 3\sigma_\epsilon^4.$$

Proof.

A typical element of $E(B_{1i}' u_i u_i' B_{1i})$ is given by

(A.C-9)

$$\begin{aligned} &E[\{Y_t(u_{t+1}-u_t)-Y_{t+1}(u_{t+2}-u_{t+1})\}\{Y_h(u_{h+1}-u_h)-Y_{h+1}(u_{h+2}-u_{h+1})\}] \\ &= E[\{Y_t(\epsilon_{t+1}-\epsilon_t)-Y_{t+1}(\epsilon_{t+2}-\epsilon_{t+1})\}\{Y_h(\epsilon_{h+1}-\epsilon_h)-Y_{h+1}(\epsilon_{h+2}-\epsilon_{h+1})\}] \\ &= E[Y_t Y_h (\epsilon_{t+1}-\epsilon_t) (\epsilon_{h+1}-\epsilon_h)] \\ &\quad - E[Y_t Y_{h+1} (\epsilon_{t+1}-\epsilon_t) (\epsilon_{h+2}-\epsilon_{h+1})] \\ &\quad - E[Y_{t+1} Y_h (\epsilon_{t+2}-\epsilon_{t+1}) (\epsilon_{h+1}-\epsilon_h)] \\ &\quad + E[Y_{t+1} Y_{h+1} (\epsilon_{t+2}-\epsilon_{t+1}) (\epsilon_{h+2}-\epsilon_{h+1})] \\ &= E(Y_t Y_h) E[(\epsilon_{t+1}-\epsilon_t) (\epsilon_{h+1}-\epsilon_h)] \\ &\quad - E(Y_t Y_{h+1}) E[(\epsilon_{t+1}-\epsilon_t) (\epsilon_{h+2}-\epsilon_{h+1})] \\ &\quad - E(Y_{t+1} Y_h) E[(\epsilon_{t+2}-\epsilon_{t+1}) (\epsilon_{h+1}-\epsilon_h)] \\ &\quad + E(Y_{t+1} Y_{h+1}) E[(\epsilon_{t+2}-\epsilon_{t+1}) (\epsilon_{h+2}-\epsilon_{h+1})] \\ &\quad + d_{th} \\ &= E(Y_t Y_h) E[(u_{t+1}-u_t) (u_{h+1}-u_h)] \\ &\quad - E(Y_t Y_{h+1}) E[(u_{t+1}-u_t) (u_{h+2}-u_{h+1})] \\ &\quad - E(Y_{t+1} Y_h) E[(u_{t+2}-u_{t+1}) (u_{h+1}-u_h)] \\ &\quad + E(Y_{t+1} Y_{h+1}) E[(u_{t+2}-u_{t+1}) (u_{h+2}-u_{h+1})] \end{aligned}$$

$$+ d_{th}$$

where $d_{th} = a_{th} - a_{t,h+1} - a_{t+1,h} + a_{t+1,h+1}$. The last equality comes from LEMMA 8. Then, we can easily show

$$(A.C-10) \quad d_{th} = \begin{cases} 2[E(\epsilon^4) - 2\sigma_\epsilon^4], & \text{if } h=t \\ -[E(\epsilon^4) - 2\sigma_\epsilon^4], & \text{if } h=t+1 \text{ or } h=t-1 \\ 0, & \text{if } h>t+1 \text{ or } h<t-1. \end{cases}$$

(A.C-9) and (A.C-10) imply the result. *QED*

In order to derive the covariance matrix of B_{1i} and B_{2i} , we need the following lemmas.

LEMMA 9

$$E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] = E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] + g_{hk},$$

$$\text{where } g_{hk} = \begin{cases} [E(\epsilon^4) - 3\sigma_\epsilon^2]/T, & \text{if } k=h \\ -[E(\epsilon^4) - 3\sigma_\epsilon^2]/T, & \text{if } k=h-1 \\ 0, & \text{otherwise.} \end{cases}$$

Proof.

If $k>h+1$ or $k<h-1$,

(A.C-11)

$$\begin{aligned} & E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] \\ &= E[\epsilon_h \{ \epsilon_{h+1} + \epsilon_h + \epsilon_{k+1} + \epsilon_k + (\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h - \epsilon_{k+1} - \epsilon_k) \} (\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h + \epsilon_{k+1} + \epsilon_k)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &\quad + E[\epsilon_h(\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h - \epsilon_{k+1} - \epsilon_k)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h + \epsilon_{k+1} + \epsilon_k)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &\quad + E[\epsilon_h(\epsilon_{k+1} + \epsilon_k)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]/T \\ &= 0 \end{aligned}$$

$$= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)].$$

If $k=h+1$,

(A.C-12)

$$\begin{aligned} E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] &= E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] \\ &= E[\epsilon_h \{ \epsilon_{h+2} + \epsilon_{h+1} + \epsilon_h + (\sum_{t=1}^T -\epsilon_{h+2} - \epsilon_{h+1} - \epsilon_h) \} (\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &= E[\epsilon_h(\epsilon_{h+2} + \epsilon_{h+1} + \epsilon_h)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &\quad + E[\epsilon_h(\sum_{t=1}^T -\epsilon_{h+2} - \epsilon_{h+1} - \epsilon_h)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &= E[\epsilon_h(\epsilon_{h+2} + \epsilon_{h+1} + \epsilon_h)(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &= E[\epsilon_h \epsilon_{h+2}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &\quad + E[\epsilon_h \epsilon_{h+1}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &\quad + E[\epsilon_h^2(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] / T \\ &= -E(\epsilon_h^2 \epsilon_{h+2}^2) / T + E(\epsilon_h^2 \epsilon_{h+1}^2) / T - E(\epsilon_h^2 \epsilon_{h+1}^2) / T \\ &= -\sigma_\epsilon^4 / T \\ &= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+2} - \epsilon_{h+1})] \\ &= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)]. \end{aligned}$$

If $k=h$,

$$\begin{aligned} \text{(A.C-13)} \quad E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] &= E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)^2] \\ &= E[\epsilon_h \{ \epsilon_{h+1} + \epsilon_h + (\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h) \} (\epsilon_{h+1} - \epsilon_h)^2] / T \\ &= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h)(\epsilon_{h+1} - \epsilon_h)^2] / T \\ &\quad + E[\epsilon_h(\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h)(\epsilon_{h+1} - \epsilon_h)^2] / T \\ &= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h)(\epsilon_{h+1} - \epsilon_h)^2] / T \\ &= E[\epsilon_h \epsilon_{h+1}(\epsilon_{h+1} - \epsilon_h)^2] / T \\ &\quad + E[\epsilon_h^2(\epsilon_{h+1} - \epsilon_h)^2] / T \\ &= -2E(\epsilon_h^2 \epsilon_{h+1}^2) / T + E(\epsilon_h^2 \epsilon_{h+1}^2) / T + E(\epsilon_h^4) / T \\ &= E(\epsilon^4) / T - \sigma_\epsilon^4 / T \end{aligned}$$

$$\begin{aligned}
&= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)^2] + [E(\epsilon^4) - 3\sigma_\epsilon^4]/T \\
&= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] + [E(\epsilon^4) - 3\sigma_\epsilon^4]/T.
\end{aligned}$$

If $k=h-1$,

(A.C-14)

$$\begin{aligned}
&E[\epsilon_h \bar{\epsilon}(\epsilon_{h+1} - \epsilon_h)(\epsilon_{k+1} - \epsilon_k)] \\
&= E[\epsilon_h \{ \epsilon_{h+1} + \epsilon_h + \epsilon_{h-1} + (\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h - \epsilon_{h-1}) \} (\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h + \epsilon_{h-1})(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&\quad + E[\epsilon_h(\sum_{t=1}^T -\epsilon_{h+1} - \epsilon_h - \epsilon_{h-1})(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&= E[\epsilon_h(\epsilon_{h+1} + \epsilon_h + \epsilon_{h-1})(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&= E[\epsilon_h \epsilon_{h+1}(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&\quad + E[\epsilon_h^2(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&\quad + E[\epsilon_h \epsilon_{h-1}(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})]/T \\
&= E(\epsilon_h^2 \epsilon_{h+1}^2)/T - E(\epsilon_h^4) + E(\epsilon_h^2 \epsilon_{h-1}^2)/T \\
&= -[E(\epsilon^4) - 2\sigma_\epsilon^4]/T \\
&= E(\epsilon_h \bar{\epsilon}) E[(\epsilon_{h+1} - \epsilon_h)(\epsilon_h - \epsilon_{h-1})] - [E(\epsilon^4) - 3\sigma_\epsilon^4]/T
\end{aligned}$$

This completes the proof. *QED*

LEMMA 10

$$E[y_t \bar{\epsilon}(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)] = E(y_t \bar{\epsilon}) E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)] + g_{hk},$$

$$\text{where } g_{hk} = \begin{cases} [E(\epsilon^4) - 3\sigma_\epsilon^2]/T, & \text{if } k=h \\ -[E(\epsilon^4) - 3\sigma_\epsilon^2]/T, & \text{if } k=h-1 \\ 0, & \text{otherwise.} \end{cases}$$

Proof.

$$\begin{aligned}
\text{(A.C-15)} \quad &E[y_t \bar{\epsilon}(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)] \\
&= \delta E(y_{t-1} \bar{\epsilon}(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)) \\
&\quad + E[(\alpha \bar{\epsilon})(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)] \\
&\quad + E[(\epsilon_t \bar{\epsilon})(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)]
\end{aligned}$$

Therefore, applying LEMMA 4,5, and 9 gives the result.

The following theorem describes the form of $E(B_{1i}'u_i u_i' B_{2i})$.

THEOREM 5

$$\begin{aligned} E(B_{1i}'u_i u_i' B_{2i}) &= \sigma_\epsilon^2 E(B_{1i}'\Gamma_i B_{2i}) + (1/T)d \cdot J_2 \\ &= \sigma_\epsilon^2 E(B_{1i}'B_{2i}) + (1/T)d \cdot J_2 \end{aligned}$$

where

$$J_2 = \begin{bmatrix} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & & & \\ 0 & -1 & 2 & & & \\ 0 & 0 & -1 & & & \\ \vdots & \vdots & \vdots & & 2 & -1 & 0 \\ 0 & 0 & 0 & & -1 & 2 & -1 \end{bmatrix}_{(T-1) \times (T-2)},$$

$$d = E(\epsilon^4) - 3\sigma_\epsilon^4.$$

Proof.

A typical element of $E(B_{1i}'u_i u_i' B_{2i})$ is given by:

$$\begin{aligned} \text{(A.C-16)} \quad E[\{Y_t(u_{t+1}-u_t) - Y_{t+1}(u_{t+2}-u_{t+1})\} \bar{u}(u_{k+1}-u_k)] \\ &= E[\{Y_t(\epsilon_{t+1}-\epsilon_t) - Y_{t+1}(\epsilon_{t+2}-\epsilon_{t+1})\} (\alpha + \bar{\epsilon})(\epsilon_{k+1}-\epsilon_k)] \\ &= E[Y_t(\alpha + \bar{\epsilon})(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\ &\quad - E[Y_{t+1}(\alpha + \bar{\epsilon})(\epsilon_{t+2}-\epsilon_{t+1})(\epsilon_{k+1}-\epsilon_k)] \end{aligned}$$

By LEMMA 5 and 10, we have

$$\begin{aligned} \text{(A.C-17)} \quad E[Y_t(\alpha + \bar{\epsilon})(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\ &= E[Y_t(\alpha + \bar{\epsilon})]E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] + g_{tk} \end{aligned}$$

$$\begin{aligned} \text{(A.C-18)} \quad E[Y_{t+1}(\alpha + \bar{\epsilon})(\epsilon_{t+2}-\epsilon_{t+1})(\epsilon_{k+1}-\epsilon_k)] \\ &= E[Y_{t+1}(\alpha + \bar{\epsilon})]E[(\epsilon_{t+2}-\epsilon_{t+1})(\epsilon_{k+1}-\epsilon_k)] + g_{t+1,k} \end{aligned}$$

Substituting (A.C-17) and (A.C-18) gives

$$\text{(A.C-19)} \quad E[\{Y_t(u_{t+1}-u_t) - Y_{t+1}(u_{t+2}-u_{t+1})\} \bar{u}(u_{k+1}-u_k)]$$

$$\begin{aligned}
&= E[Y_t(\alpha + \bar{\epsilon})]E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{k+1} - \epsilon_k)] \\
&\quad - E[Y_{t+1}(\alpha + \bar{\epsilon})]E[(\epsilon_{t+2} - \epsilon_{t+1})(\epsilon_{k+1} - \epsilon_k)] \\
&\quad + b_{tk} \\
&= E(Y_t \bar{u})E[(u_{t+1} - u_t)(u_{k+1} - u_k)] \\
&\quad - E(Y_{t+1} \bar{u})E[(u_{t+2} - u_{t+1})(u_{k+1} - u_k)] \\
&\quad + b_{tk}
\end{aligned}$$

where $b_{tk} = g_{tk} - g_{t+1,k}$. We can easily show that

$$\text{(A.C-20) } b_{tk} = \begin{cases} (2/T)d, & \text{if } k=t \\ -(1/T)d, & \text{if } k=t+1 \text{ or } k=t-1 \\ 0, & \text{otherwise.} \end{cases}$$

(A.C-19) and (A.C-20) imply the result. *QED*

the following lemmas help us derive the covariance matrix of $B_{2i}'u_i$.

LEMMA 11

$$E[\bar{\epsilon}^2(\epsilon_{t+1} - \epsilon_t)^2] = E(\bar{\epsilon}^2)E[(\epsilon_{t+1} - \epsilon_t)^2] + 2[E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2.$$

Proof.

$$\begin{aligned}
E[\bar{\epsilon}^2(\epsilon_{t+1} - \epsilon_t)^2] &= E[\bar{\epsilon}^2(\epsilon_{t+1}^2 + 2\epsilon_{t+1}\epsilon_t + \epsilon_t^2)] \\
&= E(\bar{\epsilon}^2\epsilon_{t+1}^2) - 2E(\bar{\epsilon}^2\epsilon_{t+1}\epsilon_t) + E(\bar{\epsilon}\epsilon_t^2) \\
&= [E(\epsilon^4) + (T-1)\sigma_\epsilon^4 - 4\sigma_\epsilon^4 + E(\epsilon^4) + (T-1)\epsilon_t^4]/T^2 \\
&= 2\sigma_\epsilon^2/T + 2[E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2 \\
&= E(\bar{\epsilon}^2)E[(\epsilon_{t+1} - \epsilon_t)^2] + 2[E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2 \quad \text{QED}
\end{aligned}$$

LEMMA 12

$$\begin{aligned}
E[\bar{\epsilon}^2(\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+2} - \epsilon_{t+1})] &= E(\bar{\epsilon}^2)E[(\epsilon_{t+1} - \epsilon_t)(\epsilon_{t+2} - \epsilon_{t+1})] \\
&\quad - [E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2.
\end{aligned}$$

Proof.

$$\begin{aligned}
& E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+2}-\epsilon_{t+1})] \\
&= E(\bar{\epsilon}^2\epsilon_{t+1}\epsilon_{t+2}) + E(\bar{\epsilon}^2\epsilon_t\epsilon_{t+1}) - E(\bar{\epsilon}^2\epsilon_{t+1}^2) - E(\bar{\epsilon}^2\epsilon_t\epsilon_{t+2}) \\
&= [2\sigma_\epsilon^4 + 2\sigma_\epsilon^4 - E(\epsilon^4) - (T-1)\sigma_\epsilon^4 - 2\sigma_\epsilon^4]/T^2 \\
&= -\sigma_\epsilon^4/T - [E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2 \\
&= E(\bar{\epsilon}^2)E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+2}-\epsilon_{t+1})] - [E(\epsilon^4) - 3\sigma_\epsilon^4]/T^2 \quad QED
\end{aligned}$$

LEMMA 13

$$E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+j+1}-\epsilon_{t+j})] = E(\bar{\epsilon}^2)E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+j+1}-\epsilon_{t+j})], \quad j \geq 2.$$

Proof.

$$\begin{aligned}
& E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+j+1}-\epsilon_{t+j})] \\
&= E(\bar{\epsilon}^2\epsilon_{t+1}\epsilon_{t+j+1}) + E(\bar{\epsilon}^2\epsilon_t\epsilon_{t+j}) - E(\bar{\epsilon}^2\epsilon_{t+1}\epsilon_{t+j}) - E(\bar{\epsilon}^2\epsilon_t\epsilon_{t+j+1}) \\
&= [2\sigma_\epsilon^4 + 2\sigma_\epsilon^4 - 2\sigma_\epsilon^4 - 2\sigma_\epsilon^4]/T^2 \\
&= 0 \\
&= E(\bar{\epsilon}^2)E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{t+j+1}-\epsilon_{t+j})] \quad QED
\end{aligned}$$

Now we can derive the covariance matrix of $B_{2i}'u_i$.

THEOREM 6

$$E(B_{2i}'u_i u_i' B_{2i}) = E(B_{2i}'\Gamma_i B_{2i}) + (1/T^2)dJ_2$$

where

$$J_3 = \begin{bmatrix} 2 & -1 & 0 & & \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & \ddots & \\ 0 & 0 & -1 & \ddots & \\ \vdots & \vdots & \vdots & & 2 & -1 \\ 0 & 0 & 0 & & -1 & 2 \end{bmatrix}_{(T-1) \times (T-1)},$$

$$d = E(\epsilon^4) - 3\sigma_\epsilon^4.$$

Proof.

A typical element of $E(B_{2i}'u_i u_i' B_{2i})$ takes the form of

$$\begin{aligned}
 \text{(A.C-21)} \quad E[\bar{u}^2(u_{t+1}-u_t)(u_{k+1}-u_k)] &= E[(\alpha+\bar{\epsilon})^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\
 &= E[\alpha^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\
 &\quad + 2E[\alpha\bar{\epsilon}(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\
 &\quad + E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\
 &= E(\alpha^2)E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] \\
 &\quad + E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)]
 \end{aligned}$$

By LEMMA 11,12, and 13, we have

(A.C-22)

$$E[\bar{\epsilon}^2(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] = E(\bar{\epsilon}^2)E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] + d_{tk}$$

where

$$d_{tk} = \begin{cases} 2d, & \text{if } k=t \\ -d, & \text{if } k=t+1 \text{ or } k=t-1 \\ 0, & \text{otherwise.} \end{cases}$$

Substituting (A.C-22) in (A.C-21) gives

(A.C-23)

$$\begin{aligned}
 E[\bar{u}^2(u_{t+1}-u_t)(u_{k+1}-u_k)] &= E[(\alpha+\bar{\epsilon})^2]E[(\epsilon_{t+1}-\epsilon_t)(\epsilon_{k+1}-\epsilon_k)] + d_{tk} \\
 &= E(\bar{u}^2)E[(u_{t+1}-u_t)(u_{k+1}-u_k)] + d_{tk}.
 \end{aligned}$$

(A.C-23) implies the result. *QED*

Since $H=[A \ B_1 \ B_2]$, and $H_i=[A_i \ B_{1i} \ B_{2i}]$, Theorem 1-6 can be used to derive the covariance matrix of $H'u$.

$$\text{(A.C-24)} \quad (1/N)\text{Cov}(H'u) = (1/N)E(H'uu'H)$$

$$\begin{aligned}
 &= (1/N) \sum_{i=1}^N E(H_i' u_i u_i' H_i) \\
 &= (1/N) \sum_{i=1}^N E(H_i' \Gamma_i H_i) + \sigma_\epsilon^2 J^*
 \end{aligned}$$

$$\begin{aligned}
&= \sigma_{\epsilon}^2 (1/N) \sum_{i=1}^N E(H_i' H_i) + \sigma_{\epsilon}^2 J^* \\
&= \sigma_{\epsilon}^2 (1/N) E(H' H) + \sigma_{\epsilon}^2 J^*
\end{aligned}$$

where

$$J^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & (d/\sigma_{\epsilon}^2 + \sigma_{\epsilon}^2) J_1 & (1/T) (d/\sigma_{\epsilon}^2) J_2 \\ 0 & (1/T) (d/\sigma_{\epsilon}^2) J_2 & (1/T^2) (d/\sigma_{\epsilon}^2) J_3 \end{bmatrix}$$

If ϵ is normally distributed, $d=0$. Therefore, under normality of ϵ ,

$$J^* = \begin{bmatrix} 0 & 0 & 0 \\ 0 & J & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

where $J = \sigma_{\epsilon}^2 J_1$.

APPENDIX D

Here, I show the explicit forms of Ψ_1 and Ψ_2 given in (IV.16) and (IV.17), respectively. The ϵ 's are assumed to be normally distributed. Then observe

$$\begin{aligned}
 \text{(A.D-1)} \quad \text{plim}(1/N)H'H &= \text{plim}(1/N) \sum_{i=1}^N H_i'H_i \\
 &= \text{plim}(1/N) \sum_{i=1}^N E(H_i'H_i) \\
 &= \text{plim}(1/N) \sum_{i=1}^N E \begin{bmatrix} 2y_{i0}^2 & 2y_{i0}\bar{u}_i^2 \\ 2y_{i0}\bar{u}_i & 2\bar{u}_i^2 \end{bmatrix} \\
 &= 2 \begin{bmatrix} \sigma_0^2 & \sigma_{0\alpha} \\ \sigma_{0\alpha} & \sigma_\alpha^2 + \sigma_\epsilon^2/2 \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
 \text{(A.D-2)} \quad \text{plim}(1/N)H'Y_{-1} &= \text{plim}(1/N) \sum_{i=1}^N E \begin{bmatrix} y_{i0}(y_{i1}-y_{i0}) \\ \bar{u}_i(y_{i1}-y_{i0}) \end{bmatrix} \\
 &= \begin{bmatrix} (\delta-1)\sigma_0^2 + \sigma_{0\alpha} \\ (\delta-1)\sigma_{0\alpha} + \sigma_\alpha^2 + \sigma_\epsilon^2/2 \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
 \text{(A.D-3)} \quad \text{plim}(1/N)D'Y_{-1} &= \text{plim}(1/N) \sum_{i=1}^N E \begin{bmatrix} 0 \\ (y_{i0}+y_{i1})(\epsilon_{i2}-\epsilon_{i1}) \end{bmatrix} \\
 &= \begin{bmatrix} 0 \\ -\sigma_\epsilon^2/2 \end{bmatrix}
 \end{aligned}$$

Using (A.D-1), (A.D-2), and (A.D-3), we have

$$\text{(A.D-4)} \quad \text{plim}(1/N)Y_{-1}'D(H'H)^{-1}H'Y_{-1} = -\sigma_\epsilon^2/4$$

$$\begin{aligned}
 \text{(A.D-5)} \quad \text{plim}(1/N)Y_{-1}'D(H'H)^{-1}D'Y_{-1} \\
 = (1/8)\sigma_\epsilon^4 / \{\sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2\}
 \end{aligned}$$

where $\rho_{0\alpha} = \sigma_{0\alpha}/(\sigma_0\sigma_\alpha)$. Substituting (A.D-4) and (A.D-5) into (IV.16) gives

(A.D-6)

$$\psi_1 = \text{plim}(1/N)Y_{-1}'P_{HY-1} + (1/8)\sigma_\epsilon^4 / \{\sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2\} - (1/2)\sigma_\epsilon^2$$

Also, using (A.D-1), (A.D-2), and (A.D-3), we can show

$$\begin{aligned} \text{(A.D-7)} \quad \text{plim}(1/N)Y_{-1}'M_AB_2 &= \text{plim}(1/N)Y_{-1}'B_2 \\ &\quad - \text{plim}(1/N)Y_{-1}'A(A'A)^{-1}A'B_2 \\ &= (\sigma_\alpha^2\sigma_0^2 - \sigma_{0\alpha}^2) / \sigma_0^2 + \sigma_\epsilon^2/2 \\ &= \sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2 \end{aligned}$$

$$\begin{aligned} \text{(A.D-8)} \quad \text{plim}(1/N)B_2'M_AB_2 &= \text{plim}(1/N)B_2'B_2 \\ &\quad - \text{plim}(1/N)B_2'A(A'A)^{-1}A'B_2 \\ &= 2\{(\sigma_\alpha^2\sigma_0^2 - \sigma_{0\alpha}^2) / \sigma_0^2 + \sigma_\epsilon^2/2\} \\ &= 2[\sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2] \end{aligned}$$

Therefore,

$$\begin{aligned} \text{(A.D-9)} \quad \text{plim}(1/N)Y_{-1}'M_AB_2(B_2'M_AB_2)^{-1}B_2'M_AY_{-1} \\ &= (1/2)\{(\sigma_\alpha^2\sigma_0^2 - \sigma_{0\alpha}^2) / \sigma_0^2 + \sigma_\epsilon^2/2\} \\ &= (1/2)[\sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2] \end{aligned}$$

Since $\text{plim}(1/N)Y_{-1}'P_AY_{-1} = \text{plim}(1/N)Y_{-1}'P_{HY-1}$

$$- \text{plim}(1/N)Y_{-1}'M_AB_2(B_2'M_AB_2)^{-1}B_2'M_AY_{-1},$$

$$\begin{aligned} \text{(A.D-10)} \quad \psi_2 &= \text{plim}(1/N)Y_{-1}'P_AY_{-1} = \text{plim}(1/N)Y_{-1}'P_{HY-1} \\ &\quad - (1/2)[\sigma_\alpha^2(1-\rho_{0\alpha}^2) + \sigma_\epsilon^2/2] \end{aligned}$$

APPENDIX E

Here, we show the efficiency gain of $\hat{\theta}_s$ defined in (V.16) over $\hat{\theta}_F$ in (V.13). Consider the inverse of the asymptotic covariance matrix of $\hat{\theta}_F$; $W'F(F'\Gamma F)^{-1}F'W$. For simplicity, let $E=[W^{**} Q_V X]$.

$$(A.E-1) \quad F'\Gamma F = \begin{bmatrix} E' \\ R'P_V \end{bmatrix} (Q_V + \phi^2 P_V) (E \ P_V R) = \begin{bmatrix} E'E & 0 \\ 0 & \phi^2 R'P_V R \end{bmatrix}$$

$$(A.E-2) \quad F'W = \begin{bmatrix} E' \\ R'P_V \end{bmatrix} (Y_{-1} \ X \ Z) = \begin{bmatrix} E'Y_{-1} & E'X & E'Z \\ R'P_V Y_{-1} & R'P_V X & R'P_V Z \end{bmatrix}$$

Using (A.E-1) and (A.E-2), some straightforward algebra shows

$$(A.E-3) \quad W'F(F'\Gamma F)^{-1}F'W = \begin{bmatrix} Y_{-1}'P_E Y_{-1} & Y_{-1}'P_E X & 0 \\ X'P_E Y & X'P_E X & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ + \phi^2 \begin{bmatrix} Y_{-1}'P_{P_V R} Y_{-1} & Y_{-1}'P_{P_V R} X & Y_{-1}'P_{P_V R} Z \\ X'P_{P_V R} Y & X'P_{P_V R} X & X'P_{P_V R} Z \\ Z'P_{P_V R} Y & Z'P_{P_V R} X & Z'P_{P_V R} Z \end{bmatrix}$$

Now, consider the inverse of the asymptotic covariance matrix of $\hat{\theta}_s$; $W'S(S'\Gamma S)^{-1}S'W$

$$(A.E-4) \quad S'\Gamma S = \begin{bmatrix} A' \\ E' \\ R'P_V \end{bmatrix} (Q_V + \phi^2 P_V) (A \ E \ F) = \begin{bmatrix} A'A & A'E & 0 \\ E'A & E'E & 0 \\ 0 & 0 & \phi^2 R'P_V R \end{bmatrix}$$

$$(A.E-5) \quad S'W = \begin{bmatrix} A' \\ E' \\ R'P_V \end{bmatrix} (Y_{-1} \ X \ Z) = \begin{bmatrix} A'Y_{-1} & A'X & 0 \\ E'Y & E'X & 0 \\ R'P_V Y_{-1} & R'P_V X & R'P_V Z \end{bmatrix}$$

Then, after some matrix operations, we have

(A.E-5) $W'S(S'RS)^{-1}S'W$

$$= \left[\begin{array}{cc|c} Y_{-1}'A & Y_{-1}'E & 0 \\ X'A & X'E & 0 \\ \hline 0 & 0 & 0 \end{array} \right] \begin{bmatrix} A'A & A'E \\ E'A & E'E \end{bmatrix}^{-1} \begin{bmatrix} A'Y_{-1} & A'X \\ E'Y_{-1} & E'X \end{bmatrix} \\ + \phi^2 \begin{bmatrix} Y_{-1}'P_{P_V R}Y_{-1} & Y_{-1}'P_{P_V R}X & Y_{-1}'P_{P_V R}Z \\ X'P_{P_V R}Y_{-1} & X'P_{P_V R}X & X'P_{P_V R}Z \\ Z'P_{P_V R}Y_{-1} & Z'P_{P_V R}X & Z'P_{P_V R}Z \end{bmatrix}$$

Note that

$$(A.E-7) \begin{bmatrix} A'A & A'E \\ E'A & E'E \end{bmatrix}^{-1} = \begin{bmatrix} I & \\ & -(E'E)^{-1}E'A \end{bmatrix} (A'M_E A)^{-1} (I, -A'E(E'E)^{-1}) \\ + \begin{bmatrix} 0 & 0 \\ 0 & (E'E)^{-1} \end{bmatrix}$$

$$(A.E-8) (I, -A'E(E'E)^{-1}) \begin{bmatrix} A'Y_{-1} & A'X \\ E'Y_{-1} & E'X \end{bmatrix} = (A'M_E Y_{-1}, A'M_E X) \\ = (A'M_E Y_{-1}, 0)$$

Therefore,

$$(A.E-9) \begin{bmatrix} Y_{-1}'A & Y_{-1}'E \\ X'A & X'E \end{bmatrix} \begin{bmatrix} A'A & A'E \\ E'A & E'E \end{bmatrix}^{-1} \begin{bmatrix} A'Y_{-1} & A'X \\ E'Y_{-1} & E'X \end{bmatrix} \\ = \begin{bmatrix} Y_{-1}'A & Y_{-1}'E \\ X'A & A'E \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & E(E'E)^{-1} \end{bmatrix} \begin{bmatrix} A'Y_{-1} & A'X \\ E'Y_{-1} & E'X \end{bmatrix} \\ + \begin{bmatrix} Y & 'M_E A \\ 0 & \end{bmatrix} A'M_E A (A'M_E Y_{-1}, 0) \\ = \begin{bmatrix} Y_{-1}'P_{E Y_{-1}} & Y_{-1}'P_X \\ X'P_{E Y_{-1}} & X'P_{E X} \end{bmatrix} + \begin{bmatrix} K_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

where $K_{11} = Y_{-1}'M_E A (A'M_E A)^{-1} A'M_E Y_{-1}$. Substituting (A.E-9) into

(A.E-6) gives

$$(A.E-10) \quad W'S(S'FS)^{-1}S'W = W'F(F'FF)^{-1}F'W + K_1$$

where

$$K = \begin{bmatrix} K_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

APPENDIX F

If the ϵ 's are normally distributed, the inverse of the asymptotic covariance matrix of $\hat{\theta}_L$ is given by:

$$(A.F-1) \quad W'(L+L_D)\zeta^{-1}(L+L_D)'W$$

$$\begin{aligned}
 &= \begin{bmatrix} Y_{-1}'(H+D) & Y_{-1}'E & Y_{-1}'P_V R \\ X'(H+D) & X'E & X'P_V R \\ Z'D & 0 & Z'P_V R \end{bmatrix} \begin{bmatrix} H'H+J^* & H'E & 0 \\ E'H & E'E & 0 \\ 0 & 0 & \phi^2 R'P_V R \end{bmatrix}^{-1} \\
 &\quad \cdot \begin{bmatrix} (H+D)Y_{-1} & (H+D)'X & (H+D)'X \\ E'Y_{-1} & E'X & 0 \\ R'P_V Y_{-1} & R'P_V X & R'P_V Z \end{bmatrix} \\
 &= \left[\begin{array}{cc|cc} Y_{-1}'(H+D) & Y_{-1}'E & H'H+J^* & H'E \\ X'(H+D) & X'E & E'H & E'E \end{array} \right]^{-1} \begin{bmatrix} (H+D)Y_{-1} & (H+D)'X \\ E'Y_{-1} & E'X \end{bmatrix} \\
 &\quad - \left[\begin{array}{cc|cc} Z'D & 0 & H'H+J^* & H'E \\ & & E'H & E'E \end{array} \right]^{-1} \begin{bmatrix} (H+D)Y_{-1} & (H+D)'X \\ E'Y_{-1} & E'X \end{bmatrix} \\
 &\quad - \left[\begin{array}{cc|cc} Y_{-1}'(H+D) & Y_{-1}'E & H'H+J^* & H'E \\ X'(H+D) & X'E & E'H & E'E \end{array} \right]^{-1} \begin{bmatrix} D'Z \\ 0 \end{bmatrix} \\
 &\quad - \left[\begin{array}{cc|cc} Z'D & 0 & H'H+J^* & H'E \\ & & E'H & E'E \end{array} \right]^{-1} \begin{bmatrix} D'Z \\ 0 \end{bmatrix} \\
 &\quad + \phi^2 \begin{bmatrix} Y_{-1}'P_V R Y_{-1} & Y_{-1}'P_V R X & Y_{-1}'P_V B Z \\ X'P_V B Y_{-1} & X'P_V B X & X'P_V B Z \\ Z'P_V B Y_{-1} & Z'P_V B X & Z'P_V B Z \end{bmatrix}
 \end{aligned}$$

Note that

$$\begin{aligned}
 (A.F-2) \quad \begin{bmatrix} H'H+J^* & H'E \\ E'H & E'E \end{bmatrix}^{-1} &= \begin{bmatrix} I & \\ -E(E'E)^{-1}H & \end{bmatrix} (H'P_H + J^*) (I - HE(E'E)^{-1}) \\
 &\quad + \begin{bmatrix} 0 & 0 \\ 0 & (E'E)^{-1} \end{bmatrix}
 \end{aligned}$$

$$(A.F-3) \quad \left(\begin{array}{cc} I & -H'E(E'E)^{-1} \end{array} \right) \begin{bmatrix} (H+D)Y_{-1} & (H+D)'X \\ E'Y_{-1} & E'X \end{bmatrix} = (H'M_E Y \quad D'X)$$

$$(A.F-4) \quad \left(\begin{array}{cc} I & -H'E(E'E)^{-1} \end{array} \right) \begin{bmatrix} D'Z \\ 0 \end{bmatrix} = D'Z$$

Substituting (A.F-2), (A.F-3), and (A.F-4) into (A.F-5) gives

$$\begin{aligned} (A.F-5) \quad & W'(L+L_D)\zeta^{-1}(L+L_D)'W \\ & = W'F(F'F)^{-1}F'W \\ & + \begin{bmatrix} Y_{-1}'M_E H Y_{-1}'D \\ X'D \\ Z'D \end{bmatrix} (H'M_E H + J^*) (H'M_E Y + D'Y_{-1} \quad D'X \quad D'Z) \\ & = W'F(F'F)^{-1}F'W \\ & + \begin{bmatrix} Y_{-1}'M_E A & Y_{-1}'M_E (B^* + D^*) \\ 0 & X'D^* \\ 0 & Z'D^* \end{bmatrix} \begin{bmatrix} A'M_E A & A'M_E B^* \\ B^{*'}M_E A & B^{*'}M_E B^* + \sigma_\epsilon^2 J^{**} \end{bmatrix}^{-1} \\ & \cdot \begin{bmatrix} A'M_E Y_{-1} & 0 & 0 \\ (B^* + D^*)'M_E Y_{-1} & D^{*'}X & D^{*'}Z \end{bmatrix} \end{aligned}$$

where $B^* = (B_1 \ B_2)$; $D^* = (0 \ D_2)$;

$$J^{**} = \sigma_\epsilon^2 \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix}$$

Note that

$$\begin{aligned} (A.F-6) \quad & \begin{bmatrix} A'M_E A & A'M_E B^* \\ B^{*'}M_E A & B^{*'}M_E B^* + \sigma_\epsilon^2 J^{**} \end{bmatrix}^{-1} \\ & = \begin{bmatrix} (A'M_E A)^{-1} & A'M_E B^* \\ -I & \end{bmatrix} (B^{*'}M_E B^* + J^{**} \quad -B^{*'}M_E A (A'M_E A)^{-1} A'M_E B^*) \end{aligned}$$

$$\cdot (B^*{}'M_E A(A' M_E A)^{-1} - I) + \begin{bmatrix} (A' M_E A)^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$= K_3 + \begin{bmatrix} (A' M A)^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

Substitute (A.F-6) into (A.F-5); then, we have

$$(A.F-7) \quad W'(L+L_D)\zeta^{-1}(L+L_D)'W = W'F(F'FF)^{-1}F'W + K_1 + K_2$$

$$= W'S(S'FS)^{-1}S'W + K_2$$

where

$$K_2 = \begin{bmatrix} Y_{-1}'M_E A & Y_{-1}'M_E (B^* + D^*) \\ 0 & X'D^* \\ 0 & Z'D^* \end{bmatrix} K_3 \begin{bmatrix} A'M Y & 0 & 0 \\ (B^* + D^*)'M_E Y_{-1} & D^{*'}X & D^{*'}Z \end{bmatrix}$$

which is a positive semidefinite matrix. This shows that $\hat{\theta}_L$ is more efficient than $\hat{\theta}_S$.

REFERENCES

REFERENCES

- Amemiya, T. (1974), "The Nonlinear Two-stage Least-squares Estimator," Journal of Econometrics, Vol. 2, pp 105-110.
- Amemiya, T., and MaCurdy, T. (1986), "Instrumental-variable Estimation of An Error-components Model," Econometrica, Vol. 54, pp 869-880.
- Anderson, T., and Hsiao, C. (1981), "Estimation of Dynamic Model with Error Components," Journal of the American Statistic Association, Vol 76, pp 598-606.
- Arellano, M., and Bond, S. (1988), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and Application to Employment Equation," working Paper no. 88/4, Institute for Fiscal studies, London.
- Balestra, P., and Nerlove, M. (1966), "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," Econometrica, Vol 34, pp 585-612.
- Breusch, T, Mizon, G., and Schmidt, P. (1989), "Efficient Estimation Using Panel Data," Econometrica, Vol. 47, pp 695-701.
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," Journal of Econometrics, Vol 34, pp 305-334.
- Hausman, J. (1978), "Specification Tests in Econometrics", Econometrica, Vol 46, pp 1251-1271.
- Hausman, J., and Taylor, W. (1981), "Panel Data and Unobservable Individual Effects," Econometrica, Vol 49, 1981, pp 1377-1399.
- Holtz-Eakin, D. (1988), "Testing for Individual Effects in Autoregressive Models," Journal of Econometrics, Vol. 39, pp 297-307.
- Hsiao, C. (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," Journal of Econometrics, Annals of Applied Econometrics, Vol. 18, pp 47-82
- Hsiao, C. (1986), Analysis of Panel Data, New York: Cambridge University Press.

Kiefer, N. (1980), "Estimation of Fixed Effect Models for Time Series of Cross-sections with Arbitrary Intertemporal Covariances," Journal of Econometrics, Vol. 14, pp 195-202

Newey, W. (1985), "Generalized Method of Moments Specification Testing," Journal of Econometrics, Vol. 29, pp 229-256.

MICHIGAN STATE UNIV. LIBRARIES



31293009063680