Employment by Occupation and Industry, 2004 and 2014: Technical Documentation

by

Peter Dixon and Maureen Rimmer *

May 6, 2006

0. Introduction

We create USAGE employment matrices for 2004 and 2014 showing employment by 753 occupations and the 513 USAGE industries. The starting point for these two matrices is BLS data for 2004 and forecasts for 2014 showing employment by 1022 occupations and 486 NAICS industries. The occupations are an almost complete set of those in the Standard Occupational Classification. The industries are those in NAICS (North American Industry Classification System) at roughly the 4-digit level. The BLS data and forecasts, with documentation, are available at http://www.bls.gov/emp/empocc2.htm as files pbmt.zip and matdoc.txt. The matdoc.txt file documents the file layout of the data matrices and provides a listing of the occupational and industry categories.

The creation of the USAGE employment matrices involved several steps.

Step 1 Many of the 1022 rows and 486 columns in the BLS data and forecasts refer to subtotals showing employment for groups of occupations and groups of industries. We attempt to make maximum use of these subtotals for filling in missing data for primitive occupations and industries. At the end of this process, we have 753 primitive occupations and 324 primitive industries.

Step 2 To a large extent the USAGE industries are those identified in the BEA benchmark input-output tables for 1992. Thus, to use the BLS data and forecasts in USAGE, it is necessary to create a concordance between BEA input-output industry classifications and NAICS classifications. To this end, we create a 513 by 324 concordance matrix of 1's and 0's. A "1" in the (i,j)-th position implies that USAGE industry i is partly contained in NAICS industry j or equivalently NAICS industry j is partly contained in

---

* We thank Kyle Johnson for a major contribution to the work reported in section 2.

USAGE industry i: that is, there is a non-zero intersection between the activities in USAGE industry i and those in NAICS industry j.

Step 3 The BLS matrices refer to employees. However, the BLS also gives a vector for 2004 and for 2014 for self-employed workers and unpaid family helpers classified by occupation. We allocate these vectors across the 324 primitive NAICS industries.

Step 4 Using (a) the 513 by 324 concordance matrix and (b) a 513 USAGE wagebill vector, we make preliminary estimates for each of the 324 NAICS industries of the shares of their employment that are accounted for by each of the 513 USAGE industries. We denote this preliminary matrix by F_1. Then we compute preliminary estimates of the matrices of employment by 753 occupations and 513 USAGE industries as

$$EU_1(o,u,t) = \sum_n EN(o,n,t) * F_1(u,n) \quad (0.1)$$

where $EU_1(o,u,t)$ is the preliminary or first estimate of employment in occupation o and USAGE industry u in year t (t= 2004, 2014) and EN(o,n,t) is employment in occupation o and NAICS industry n in year t obtained from the BLS data and forecasts.

Step 5 The employment matrices (EU_1) created in step 4 together with USAGE wagebill data allow us to compute implied wage rates for the 513 USAGE industries. Some of these implied wage rates are implausible. We make two successive modifications of the F matrix in a way that leads to more plausible implied industry wage rates without violating our USAGE-NAICS concordance. In this process we create EU_2 and EU_3.

Step 6 For each occupation we look at the distribution of employment across USAGE industries implied by EU_3. For some occupations this is clearly unrealistic. This leads us to create F matrices with an occupational dimension. That is we create matrices FO(o,u,n) that show, for each of the 324 NAICS industries, the shares of their occupation-o employment that are accounted for by each of the 513 USAGE industries. Then we compute our final matrices of employment by occupation and USAGE industry for year t as:

$$EU_{final}(o,u,t) = \sum_n EN(o,n,t) * FO(o,u,n) \quad (0.2)$$

Sections 1 to 6 of this paper contain details of these steps in the estimation of the USAGE employment matrix for 2004. Section 7 describes the estimation of the USAGE employment matrix for 2014. Section 8 compares the USAGE employment matrices for 2004 and 2014. Section 9 describes how these matrices are used in the USAGE model.

1. Step 1: creating primitive employment matrix for 2004

Column (2) of Table 1.1 lists most of the 486 industry classifications included in the BLS employment matrices. Only a few classifications, referring to non-industry-specific aggregates are omitted. Column (1) indicates whether the BLS considers the classification to be a line item or a summary. For the most part, summaries are aggregations of line items. There are four exceptions: classifications 443000, 946110, 946220 and 949400. These are identified by the BLS as summary items. However, for these classifications the BLS data and forecasts contain no underlying line items. Consequently, we treat these four classifications as if they were line items. With these four classifications included, the BLS data contains 324 industry line items.

Each row of columns (3) to (8) of Table 1.1 contains one entry. This entry refers to the NAICS classification in column (2). For example, n113300 refers NAICS industry 113300; n113310 refers to NAICS industry 113310; n238161 refers to NAICS industry 238120,30,50,70,90; and n31X330 refers to NAICS industry 31-33. The name changes (including the addition of "n", the elimination hyphens and commas and the shortening of names) as we go from column (2) to columns (3) - (8) were undertaken to facilitate computer manipulations and to avoid confusions with USAGE industrial classifications.

The positioning of the entries in columns (3) to (8) of Table 1.1 defines the hierarchies in the BLS NAICS-based industrial classifications. Entries in column (4) are subcategories of the immediately preceding entry in column (3). For example, n236000, n237000 and n238000 are a subcategories of n230000. Entries in column (5) are subcategories of the immediately preceding entry in column (4). For example, n238100, n238200, n238300 and n238900 are a subcategories of n238000. Entries in column (6) are subcategories of the immediately preceding entry in column (5). For example, n238110, n238140, n238160 and n238161 are a subcategories of n238100. Similarly, entries in column (7) are subcategories of the immediately preceding entry in column (6), and entries in column (8) are subcategories of the immediately preceding entry in column (7).

We could have drawn up an occupations table, similar to Table 1.1, but showing hierarchies in the BLS occupational classifications. However, as we will see, industry-classification hierarchies play a role in our data manipulations but occupational-classification hierarchies do not. For our data manipulations, we concentrate only on the line item

Table 1.1. NAICS industrial classifications in the BLS employment matrices

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Summary	113300	n113300					
Line item	113310		n113310				
Summary	210000	n210000					
Summary	211000		n211000				
Line item	211100			n211100			
Summary	212000		n212000				
Line item	212100			n212100			
Line item	212200			n212200			
Line item	212300			n212300			
Summary	213000		n213000				
Line item	213100			n213100			
Summary	220000	n220000					
Summary	221000		n221000				
Line item	221100			n221100			
Line item	221200			n221200			
Line item	221300			n221300			
Summary	230000	n230000					
Summary	236000		n236000				
Line item	236100			n236100			
Line item	236200			n236200			
Summary	237000		n237000				
Summary	237100			n237100			
Line item	237130				n237130		
Line item	237110-20				n237131		
Line item	237200			n237200			
Line item	237300			n237300			
Line item	237900			n237900			
Summary	238000		n238000				
Summary	238100			n238100			
Line item	238110				n238110		
Line item	238140				n238140		
Line item	238160				n238160		
Line item	238120,30,50,70,90				n238161		
Summary	238200			n238200			
Line item	238210				n238210		
Line item	238220				n238220		
Line item	238290				n238290		
Summary	238300			n238300			
Line item	238310				n238310		
Line item	238320				n238320		
Line item	238330-50,90				n238321		
Line item	238900			n238900			
Summary	31-33	n31X330					
Summary	311000		n311000				
Line item	311100			n311100			
Line item	311200			n311200			
Line item	311300			n311300			
Line item	311400			n311400			
Line item	311500			n311500			
Line item	311600			n311600			
Line item	311700			n311700			
Line item	311800			n311800			
Line item	311900			n311900			
Summary	312000		n312000				
Line item	312100			n312100			
Line item	312200			n312200			
Summary	313000		n313000				
Line item	313100			n313100			
Line item	313200			n313200			
Line item	313300			n313300			
Summary	314000		n314000				
Line item	314100			n314100			
Line item	314900			n314900			
Summary	315000		n315000				
Line item	315100			n315100			
Line item	315200			n315200			
Line item	315900			n315900			
Summary	316000		n316000				
Line item	316200			n316200			
Summary	321000		n321000				
Line item	321100			n321100			
Line item	321200			n321200			
Line item	321900			n321900			
Summary	322000		n322000				
Line item	322100			n322100			
Line item	322200			n322200			

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Summary	323000		n323000				
Line item	323100			n323100			
Summary	324000		n324000				
Line item	324100			n324100			
Summary	325000		n325000				
Line item	325100			n325100			
Line item	325200			n325200			
Line item	325300			n325300			
Line item	325400			n325400			
Line item	325500			n325500			
Line item	325600			n325600			
Line item	325900			n325900			
Summary	326000		n326000				
Line item	326100			n326100			
Line item	326200			n326200			
Summary	327000		n327000				
Line item	327100			n327100			
Line item	327200			n327200			
Line item	327300			n327300			
Summary	331000		n331000				
Line item	331100			n331100			
Line item	331200			n331200			
Line item	331300			n331300			
Line item	331400			n331400			
Line item	331500			n331500			
Summary	332000		n332000				
Line item	332100			n332100			
Line item	332200			n332200			
Line item	332300			n332300			
Line item	332400			n332400			
Line item	332500			n332500			
Line item	332600			n332600			
Summary	332700			n332700			
Line item	332710				n332710		
Line item	332720				n332720		
Line item	332800			n332800			
Line item	332900			n332900			
Summary	333000		n333000				
Line item	333100			n333100			
Line item	333200			n333200			
Line item	333300			n333300			
Line item	333400			n333400			
Line item	333500			n333500			
Line item	333600			n333600			
Line item	333900			n333900			
Summary	334000		n334000				
Line item	334100			n334100			
Line item	334200			n334200			
Line item	334300			n334300			
Line item	334400			n334400			
Line item	334500			n334500			
Line item	334600			n334600			
Summary	335000		n335000				
Line item	335100			n335100			
Line item	335200			n335200			
Line item	335300			n335300			
Line item	335900			n335900			
Summary	336000		n336000				
Line item	336100			n336100			
Line item	336200			n336200			
Line item	336300			n336300			
Line item	336400			n336400			
Line item	336500			n336500			
Line item	336600			n336600			
Line item	336900			n336900			
Summary	337000		n337000				
Summary	337100			n337100			
Line item	337110				n337110		
Line item	337120				n337120		
Line item	337200			n337200			
Line item	337900			n337900			
Summary	339000		n339000				
Line item	339100			n339100			
Summary	339900			n339900			
Line item	339910				n339910		
Line item	339920-50,90				n339920		
Summary	420000	n420000					
Summary	423000		n423000				

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Line item	423100			n423100			
Line item	423200			n423200			
Line item	423300			n423300			
Line item	423400			n423400			
Line item	423500			n423500			
Line item	423600			n423600			
Line item	423700			n423700			
Summary	423800			n423800			
Line item	423820				n423820		
Line item	423810,30-60				n423830		
Line item	423900			n423900			
Summary	424000		n424000				
Line item	424100			n424100			
Line item	424200			n424200			
Line item	424300			n424300			
Line item	424400			n424400			
Line item	424500			n424500			
Line item	424600			n424600			
Line item	424700			n424700			
Line item	424800			n424800			
Line item	424900			n424900			
Summary	425000		n425000				
Line item	425100			n425100			
Summary	44-45	n44X450					
Summary	441000		n441000				
Line item	441100			n441100			
Line item	441200			n441200			
Line item	441300			n441300			
Summary	442000		n442000				
Line item	442100			n442100			
Line item	442200			n442200			
Summary	443000		n443000				
Summary	444000		n444000				
Line item	444100			n444100			
Line item	444200			n444200			
Summary	445000		n445000				
Line item	445100			n445100			
Line item	445200			n445200			
Line item	445300			n445300			
Summary	446000		n446000				
Summary	446100			n446100			
Line item	446110				n446110		
Line item	446120,30,90				n446120		
Summary	447000		n447000				
Line item	447100			n447100			
Summary	448000		n448000				
Line item	448100			n448100			
Line item	448200			n448200			
Line item	448300			n448300			
Summary	451000		n451000				
Summary	451100			n451100			
Line item	451110				n451110		
Line item	451120-40				n451120		
Line item	451200			n451200			
Summary	452000		n452000				
Line item	452100			n452100			
Line item	452900			n452900			
Summary	453000		n453000				
Line item	453100			n453100			
Line item	453200			n453200			
Line item	453300			n453300			
Line item	453900			n453900			
Summary	454000		n454000				
Line item	454100			n454100			
Line item	454200			n454200			
Line item	454300			n454300			
Summary	48-49	n48X490					
Summary	481000		n481000				
Line item	481100			n481100			
Line item	481200			n481200			
Summary	482000		n482000				
Line item	482100			n482100			
Summary	483000		n483000				
Line item	483100			n483100			
Line item	483200			n483200			
Summary	484000		n484000				
Line item	484100			n484100			

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Line item	484200			n484200			
Summary	485000		n485000				
Line item	485100			n485100			
Line item	485200			n485200			
Line item	485300			n485300			
Line item	485400			n485400			
Line item	485500			n485500			
Line item	485900			n485900			
Summary	488000		n488000				
Line item	488100			n488100			
Line item	488300			n488300			
Line item	488400			n488400			
Line item	488500			n488500			
Summary	492000		n492000				
Line item	492100			n492100			
Line item	492200			n492200			
Summary	493000		n493000				
Line item	493100			n493100			
Summary	510000	n510000					
Summary	511000		n511000				
Summary	511100			n511100			
Line item	511110				n511110		
Line item	511120-40,90				n511120		
Line item	511200			n511200			
Summary	512000		n512000				
Summary	512100			n512100			
Line item	512130				n512130		
Line item	512110,20,90				n512140		
Line item	512200			n512200			
Summary	515000		n515000				
Summary	515100			n515100			
Line item	515110				n515110		
Line item	515120				n515120		
Line item	515200			n515200			
Summary	517000		n517000				
Line item	517100			n517100			
Line item	517200			n517200			
Line item	517300			n517300			
Line item	517500			n517500			
Summary	516,8,9		n516890				
Summary	516000			n516000			
Line item	516100				n516100		
Summary	518000			n518000			
Line item	518100				n518100		
Line item	518200				n518200		
Summary	519000			n519000			
Line item	519100				n519100		
Summary	520000	n520000					
Summary	521-2		n521X20				
Summary	521000			n521000			
Line item	521100				n521100		
Summary	522000			n522000			
Line item	522100				n522100		
Summary	522200				n522200		
Line item	522290					n522290	
Line item	522210-20					n522295	
Line item	522300				n522300		
Summary	523000		n523000				
Line item	523900			n523900			
Summary	524000		n524000				
Summary	524100			n524100			
Line item	524120				n524120		
Line item	524110,30				n524130		
Summary	524200			n524200			
Line item	524210				n524210		
Line item	524290				n524290		
Summary	525000		n525000				
Line item	525100			n525100			
Line item	525900			n525900			
Summary	530000	n530000					
Summary	531000		n531000				
Line item	531100			n531100			
Line item	531200			n531200			
Line item	531300			n531300			
Summary	532000		n532000				
Line item	532100			n532100			
Summary	5322-3			n5322X3			

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Line item	532200				n532200		
Line item	532300				n532300		
Line item	532400			n532400			
Summary	533000		n533000				
Line item	533100			n533100			
Summary	540000	n540000					
Summary	541000		n541000				
Line item	541100			n541100			
Line item	541200			n541200			
Summary	541300			n541300			
Line item	541310-20				n541312		
Line item	541330-70				n541337		
Line item	541380				n541380		
Line item	541400			n541400			
Line item	541500			n541500			
Line item	541600			n541600			
Summary	541700			n541700			
Line item	541710				n541710		
Line item	541720				n541720		
Line item	541800			n541800			
Summary	541900			n541900			
Line item	541920				n541920		
Line item	541940				n541940		
Line item	541910,30,90				n541930		
Summary	550000	n550000					
Summary	551000		n551000				
Line item	551100			n551100			
Summary	560000	n560000					
Summary	561000		n561000				
Line item	561100			n561100			
Line item	561200			n561200			
Line item	561300			n561300			
Summary	561400			n561400			
Line item	561420				n561420		
Line item	561410, 30-90				n561430		
Summary	561500			n561500			
Line item	561510				n561510		
Line item	561520-90				n561520		
Summary	561600			n561600			
Line item	561610				n561610		
Line item	561620				n561620		
Summary	561700			n561700			
Line item	561710				n561710		
Line item	561730				n561730		
Line item	561720, 40-90				n561740		
Line item	561900			n561900			
Summary	562000		n562000				
Line item	562100			n562100			
Summary	562200-900			n5622X9			
Line item	562200				n562200		
Line item	562900				n562900		
Summary	610000	n610000					
Summary	611000		n611000				
Line item	611100			n611100			
Summary	611200-300			n6112X3			
Line item	611200				n611200		
Line item	611300				n611300		
Summary	611400-700			n6114X7			
Line item	611400				n611400		
Line item	611500				n611500		
Line item	611600				n611600		
Line item	611700				n611700		
Summary	620000	n620000					
Summary	621000-3000		n621X30				
Summary	621000			n621000			
Summary	621100-300				n6211X3		
Line item	621100					n621100	
Line item	621200					n621200	
Summary	621300					n621300	
Line item	621310						n621310
Line item	621320						n621320
Line item	621330						n621330
Line item	621340						n621340
Line item	621390						n621390
Line item	621600				n621600		
Summary	621400, 500, 900				n621459		
Summary	621400					n621400	

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1st tier (3)	2nd tier (4)	3rd tier (5)	4th tier (6)	5th tier (7)	6th tier (8)
Line item	621420						n621420
Line item	621410, 90						n621430
Line item	621500					n621500	
Summary	621900					n621900	
Line item	621910						n621910
Line item	621990						n621990
Summary	622000			n622000			
Line item	622100				n622100		
Line item	622200				n622200		
Line item	622300				n622300		
Summary	623000			n623000			
Line item	623100				n623100		
Summary	623200-900				n6232X9		
Summary	623200					n623200	
Line item	623210						n623210
Line item	623220						n623220
Line item	623300					n623300	
Line item	623900					n623900	
Summary	624000		n624000				
Summary	624100-300			n6241X3			
Summary	624100				n624100		
Line item	624120					n624120	
Line item	624110, 90					n624130	
Line item	624200				n624200		
Line item	624300				n624300		
Line item	624400			n624400			
Summary	710000	n710000					
Summary	711000		n711000				
Line item	711200			n711200			
Line item	711500			n711500			
Summary	712000		n712000				
Line item	712100			n712100			
Summary	713000		n713000				
Line item	713100			n713100			
Line item	713200			n713200			
Summary	713900			n713900			
Line item	713940				n713940		
Line item	713910-30, 50-90				n713950		
Summary	720000	n720000					
Summary	721000		n721000				
Summary	721100, 721300			n721113			
Line item	721120				n721120		
Line item	721110, 721190, 721300				n721130		
Line item	721200			n721200			
Summary	722000		n722000				
Line item	722100			n722100			
Line item	722200			n722200			
Line item	722300			n722300			
Line item	722400			n722400			
Summary	810000	n810000					
Summary	811000		n811000				
Summary	811100			n811100			
Line item	811110				n811110		
Line item	811120				n811120		
Line item	811190				n811190		
Line item	811200			n811200			
Line item	811300			n811300			
Line item	811400			n811400			
Summary	812000		n812000				
Line item	812100			n812100			
Line item	812200			n812200			
Line item	812300			n812300			
Line item	812900			n812900			
Summary	813000		n813000				
Line item	813100			n813100			
Summary	813200-300			n8132X3			
Line item	813200				n813200		
Line item	813300				n813300		
Summary	813400-900			n8134X9			
Line item	813400				n813400		
Summary	813900				n813900		
Line item	813930					n813930	
Line item	813910-20, 40-90					n813940	
Summary	990000	n990000					
Summary	910000		n910000				
Line item	491100			n491100			
Line item	919999			n919999			

Table 1.1 continues …

Table 1.1 continued

Type (1)	Classification (2)	1^{st} tier (3)	2^{nd} tier (4)	3^{rd} tier (5)	4^{th} tier (6)	5^{th} tier (7)	6^{th} tier (8)
Summary	940000		n940000				
Summary	946110			n946110			
Summary	946220			n946220			
Summary	949400			n949400			
Summary	110000	n110000					
Line item	111000		n111000				
Line item	112000		n112000				
Line item	113100-3200		n113132				
Line item	114000		n114000				
Line item	115000		n115000				
Summary	814000	n814000					
Line item	814100		n814100				
Line item	814102		n814102				

occupational classifications of which there are 753. We ignore the summary occupational classifications. At the end of this section we return to the issue of the asymmetrical treatment of the occupational and industry hierarchies.

Having identified the 324 effective industry line items and the 753 occupational line items in the BLS data, we formed a 753 by 324 employment matrix for 2004 in which the rows and columns refer only to line items. The sum of the entries in this matrix was only 128.4 million employees. However, the sum of the entries in the BLS data for 2004 in the 1st tier industries [those identified in column (3) of Table 1.1] is 133.2 million employees. In the transition from the highly aggregated 1^{st} tier industries to the line items, 4.8 million jobs are lost. There are two reasons for this: the BLS does not publish cells containing less than 100 jobs and the BLS suppresses some cells for confidentiality reasons.

To recover the 4.8 million jobs, we proceeded in a sequential fashion.

First, we inflated the tier-2 items so that they added up to their corresponding tier-1 item. For example, we inflated occupation-o employment in n236000, n237000 and n238000 to equal occupation-o employment in n230000. In making these inflations, we started by estimating occupation-o employment in the tier-1 industries for which there are tier-2 industries as

$$S_{12}(o,j) = \sum_{k:j=M_{12}(k)} E_2(o,k) \tag{1.1}$$

for each occupation o and each tier-1 industry j with underlying tier-2 industries, where

$E_2(o,k)$ is employment in occupation o in tier-2 industry k; and
$k{:}j=M_{12}(k)$ identifies the set of tier-2 industries that are subcategories of tier-1 industry j.

Next we calculated

$$R_1(o,j) = E_1(o,j) - S_{12}(o,j) \tag{1.2}$$

where

$E_1(o,j)$ is employment in occupation o in tier-1 industry j; and
$R_1(o,j)$ is the excess of employment in occupation o in tier-1 industry j over the level estimated that on the basis of the tier-2 industries included in tier-1 industry j.

All of the $R_1(o,j)$'s are greater than or equal to zero, indicating that as we go to more detailed industry employment categories, some employment is lost through the BLS's rule concerning the minimum entry level of 100 jobs and through confidentiality. We distributed the $R_1(o,j)$'s to the tier-2 industries within tier-1 industry j as follows. If all of the tier-2 industries k within j are shown with non-zero employment in occupation o, then we reset occupation-o employment in industry k as $E_2^1(o,k)$ according to

$$E_2^1(o,k) = E_2(o,k) + R_1(o,j) * \frac{E_2(o,k)}{S_{12}(o,j)} \quad . \tag{1.3}$$

If one or more of the tier-2 industries k within j are shown with zero employment in occupation o, then we reset occupation-o employment in industry k as $E_2^1(o,k)$ according to

$$E_2^1(o,k) = E_2(o,k) + R_1(o,j) * \frac{\text{index}(o,k) * E_2(k)}{\sum_{h:j=M_{12}(h)} \text{index}(o,h) * E_2(h)} \quad . \tag{1.4}$$

where

index(o,k) is zero if $E_2(o,k) > 0$ and 1 otherwise; and
$E_2(k)$ is total employment in industry k.

If there is a 5 per cent short-fall in the occupation-o employment shown for the tier-2 component industries of tier-1 industry j [$R_1(o,j)=0.05*S_{12}(o,j)$] and all of these tier-2 industries have non-zero occupation-o employment, then under equation (1.3) we adjust the $E_2(o,k)$'s up by 5 per cent. If $E_2(o,k)$ is zero for some of the k's, then we assume that any shortfall [$R_1(o,j)$] is caused by omission of data in these zero positions. Under equation (1.4) we distribute $R_1(o,j)$ to these zero positions in proportion to employment in the relevant industries k.

Having inflated the tier-2 items so that they are consistent with the tier-1 items, we inflated the tier-3 items so that they added up to the inflated values of the corresponding tier-2 items. For example, we inflated occupation-o employment in n236100, and n236200 to equal inflated occupation-o employment in n236000. In making the tier-3 inflations, we applied formulas similar to those used in making the tier-2 inflations.

We continued this process, eventually inflating tier-6 items so that they are consistent with inflated tier-5 items.

At this stage, all of the 4.8 million missing jobs have been allocated to the 324 line item industries and the 753 line item occupations. Thus we are able to form a 753 by 324 employment matrix for 2004 in which the sum of employment is 133.2 million. This matrix has the property for each of the 753 occupations that subtotals formed as sums across subgroups of industries are consistent with our best estimates for these subtotals.

However, the matrix does not have the symmetric property. It is not true for each of the 324 industries that subtotals formed as sums across subgroups of occupations are consistent with our best estimates for these subtotals. We don't think that this is a serious problem and it could be attended to by RAS-like procedures in future research. We obtained evidence that it is not a serious problem by aggregating our matrix to 10 tier-1 occupations by 22 tier-1 industries. We then compared the 220 items in this aggregated matrix with the corresponding published BLS summary subtotals. While there are gaps between the two sets of estimates, they are relatively small. This is illustrated in Table 1.2. The rows in Table 1.2 marked with "a" are calculated from our aggregated matrix while those without "a" are calculated from the BLS data.

2. Step 2: creating a concordance between USAGE industry classifications and NAICS classifications

An essential task in going from a matrix based on NAICS industries to one, such as the USAGE classification, based on BEA input-output industries is to define a correspondence or concordance between the two classification systems. No such ready-made concordance could be found, but we do have a concordance between NAICS industries and SIC industries, and between SIC and BEA industries.

The NAICS/SIC concordance is at http://www.census.gov/epcd/naics/naicstb1.pdf, or (less legibly but more usefully) at http://www.census.gov/epcd/naics/naicstb1.csv. Part of this concordance is reproduced in Table 2.1.[1]

[1] At the Centre of Policy Studies, the full concordance can be found as sheet 1 ("sicnaics") on our Excel workbook e:\usage\kyle\jan19\sicnaicsbea4.xls, where it has been sorted in SIC order.

Table 1.2. Percentages of jobs in tier-1 occupations accounted for by tier-1 industries: comparison of estimates for 2004

Occupations \ Industries	n110000	n113300	n210000	n220000	n230000	n31X330	n420000	n44X450	n48X490	n510000	n520000	n530000	n540000	n550000	n560000	n610000	n620000	n710000	n720000	n810000	n814000	n990000
o111300	1.78	0.02	0.40	0.60	4.52	9.83	4.50	4.79	1.78	3.08	14.10	2.47	10.54	4.48	3.86	2.33	5.93	0.85	3.02	3.99	0.02	17.12
o111300a	1.78	0.01	0.40	0.60	4.52	9.84	4.50	4.79	1.78	3.08	14.11	2.46	10.54	4.48	3.86	2.32	5.93	0.84	3.01	3.99	0.02	17.13
o152900	0.05	0.00	0.19	0.31	0.35	5.11	1.18	2.17	0.23	3.82	1.75	0.19	11.20	1.25	1.88	6.33	22.65	0.75	0.10	4.36	0.02	36.10
o152900a	0.04	0.00	0.19	0.31	0.34	5.11	1.18	2.16	0.22	3.81	1.75	0.18	11.21	1.25	1.90	6.32	22.67	0.75	0.09	4.35	0.02	36.15
o313900	0.19	0.00	0.01	0.03	0.22	0.63	0.17	3.19	0.77	0.47	0.14	1.15	0.62	0.28	9.58	0.98	18.07	4.24	35.35	4.80	3.19	15.93
o313900a	0.19	0.00	0.01	0.03	0.21	0.62	0.17	3.19	0.77	0.47	0.14	1.15	0.62	0.28	9.58	0.97	18.08	4.24	35.36	4.80	3.19	15.94
o410000	0.08	0.00	0.05	0.06	1.01	3.09	10.68	59.26	0.61	2.88	5.23	3.57	2.16	0.71	3.62	0.16	0.35	1.07	2.80	1.78	0.01	0.80
o410000a	0.08	0.00	0.04	0.06	1.01	3.09	10.68	59.28	0.61	2.88	5.24	3.57	2.16	0.71	3.62	0.15	0.35	1.07	2.80	1.78	0.01	0.79
o430000	0.21	0.02	0.21	0.54	2.85	6.02	5.63	10.53	3.58	3.21	13.01	2.03	7.63	2.28	7.43	1.94	10.35	0.73	1.72	3.91	0.03	16.16
o430000a	0.21	0.02	0.20	0.54	2.84	6.02	5.63	10.53	3.58	3.21	13.01	2.03	7.63	2.28	7.43	1.94	10.35	0.73	1.71	3.91	0.03	16.16
o450000	73.06	4.28	0.05	0.04	0.10	3.95	5.18	2.86	0.35	0.00	0.02	0.16	0.47	0.53	4.31	0.11	0.06	0.63	0.08	0.43	0.04	3.28
o450000a	73.52	4.18	0.04	0.02	0.09	3.92	5.19	2.87	0.35	0.00	0.00	0.14	0.43	0.51	4.46	0.08	0.04	0.55	0.07	0.45	0.03	3.05
o470000	0.11	0.01	2.99	0.58	74.12	4.37	0.39	0.88	0.46	0.06	0.02	0.56	0.92	0.21	4.86	0.21	0.25	0.15	0.06	0.28	0.04	8.48
o470000a	0.11	0.01	2.99	0.57	74.23	4.37	0.38	0.87	0.45	0.06	0.02	0.55	0.91	0.20	4.85	0.20	0.24	0.14	0.06	0.28	0.04	8.49
o490000	0.37	0.06	0.81	2.82	8.87	13.72	7.43	15.33	5.44	5.81	0.25	5.54	1.19	0.68	3.41	0.70	2.01	1.25	1.67	12.50	0.01	10.13
o490000a	0.37	0.05	0.80	2.81	8.89	13.73	7.45	15.35	5.44	5.81	0.24	5.54	1.18	0.67	3.40	0.70	2.00	1.24	1.66	12.52	0.01	10.13
o510000	0.17	0.04	0.53	0.72	0.92	72.58	3.49	4.48	0.70	0.94	0.04	0.13	1.03	0.35	6.33	0.06	1.10	0.05	0.84	3.54	0.03	1.94
o510000a	0.17	0.03	0.52	0.71	0.92	72.71	3.49	4.48	0.68	0.96	0.03	0.12	1.02	0.33	6.33	0.05	1.09	0.04	0.84	3.53	0.03	1.93
o530000	0.53	0.14	0.84	0.14	2.72	14.29	12.19	12.89	25.54	0.91	0.05	1.49	0.65	0.59	11.56	0.27	1.27	0.43	2.28	3.76	0.07	7.40
o530000a	0.53	0.13	0.84	0.14	2.71	14.30	12.19	12.90	25.56	0.90	0.04	1.48	0.64	0.59	11.56	0.26	1.27	0.43	2.28	3.75	0.07	7.41

The concordance between SIC and BEA was constructed at the USITC for an earlier project. That correspondence is directly derived from Appendix A of the 1992 Benchmark I-O accounts of the United States.[2] Part of the correspondence is reproduced in Table 2.2.[3]

From the SIC/NAICS and SIC/BEA concordances we formed a BEA/NAICS concordance, part of which is shown in the first two columns of Table 2.3.[4] A BEA and NAICS pair appear in these first two columns only if the pair are concorded with a common SIC. However, not all BEA/NAICS pairs with a common SIC are included. We excluded pairs where we judged that the overlap was likely to be minor. To clarify the process by which the first two columns of Table 2.3 were constructed, we work through an example. From Tables 2.2 and 2.1, we can extract the information displayed in Table 2.4. This shows that the following BEA/NAICS pairs are potential inclusions in the first two columns of Table 2.3: (010100,112111); (010100,11212); (010100,111998); (010100,11239); and (010100,11299). However, using the descriptions of the industries, we made a judgment that only the pair (010100,11212) was likely to have a significant overlap of activities. Consequently out of the five potential pairs, only this one appears in Table 2.3.

The NAICS categories in column 2 of Table 2.3 are generally at a higher level of disaggregation that those in the BLS employment matrices. For each of these NAICS categories, column 3 of Table 2.3 shows the BLS/NAICS category to which it belongs.

On the basis of columns 1 and 3 of Table 2.3, we created a 513 by 324 concordance matrix of "1"s and "0"s. The 513 rows represent the USAGE industries, which in turn correspond very closely to the BEA input-output industries. The 324 columns represent the

---

2 U.S. Department of Commerce, Bureau of Economic Analysis. *Benchmark Input-Output Accounts of the United States, 1992*. Washington, DC: U.S. Government Printing Office, September 1998.

3 At the Centre of Policy Studies, the full concordance can be found on our Excel workbook e:\usage\kyle\1987SIC to 1992I-O concordance.xls.

4 At the Centre of Policy Studies, the full version of Table 2.4 can be found in the "BEAOccNAICS sorted" sheet of our Excel workbook e:\usage\kyle\jan31\sicnaicsbeaJ30.xls.

Table 2.1. A sample of the SIC/NAICS concordance

SIC		SIC description	NAICS code	NAICS description
0111		Wheat	11114	Wheat Farming
0112		Rice	11116	Rice Farming
0115		Corn	11115	Corn Farming (pt)
0116		Soybeans	11111	Soybean Farming
0119		Cash Grains, NEC		
0119	pt	Dry Pea and Bean Farms	11113	Dry Pea and Bean Farming
0119	pt	. Oilseed, Except Soybean, Farms	11112	Oilseed (except Soybean) Farming
0119	pt	. Popcorn Farms	11115	Corn Farming (pt)
0119	pt	. Combination Oilseed and Grain Farms	111191	Oilseed and Grain Combination Farming
0119	pt	. Other Farms	111199	All Other Grain Farming
0133		Sugarcane and Sugar Beets		
0133	pt	. Sugar Beets	111991	Sugar Beet Farming
0133	pt	. Sugarcane	11193	Sugarcane Farming
0134		Irish Potatoes	111211	Potato Farming
0139		Field Crops, Except Cash Grains, NEC		
0139	pt	. Hay Farms	11194	Hay Farming
0139	pt	. Peanut Farming	111992	Peanut Farming
0139	pt	. Sweet Potatoes and Yam Farms	111219	Other Vegetable (except Potato) and Melon Farming (pt)
0139	pt	. Other Field Crop Farms	111998	All Other Miscellaneous Crop Farming (pt)
0191		General Farms, Primarily Crop	111998	All Other Miscellaneous Crop Farming (pt)
0241		Dairy farms		
0241	pt	. Dairy Heifer Replacement Farms	112111	Beef Cattle Ranching and Farming (pt)
0241	pt	. Dairy Farms	11212	Dairy Cattle and Milk Production
0251		Broiler, Fryers, and Roaster Chickens	11232	Broilers and Other Meat Type Chicken Production
0252		Chicken Eggs	11231	Chicken Egg Production
0253		Turkey and Turkey Eggs	11233	Turkey Production
0254		Poultry Hatcheries	11234	Poultry Hatcheries
0259		Poultry and Eggs, NEC	11239	Other Poultry Production
0291		General Farms, Primarily Livestock & Animal Specialties	11299	All Other Animal Production (pt)

Table 2.2. A sample of the BEA/SIC concordance

BEA 1992 ID	BEA 1992 input-output sector name	SIC 1987 ID
010100	Dairy farm products	0241
010100	Dairy farm products	0191
010100	Dairy farm products	0259
010100	Dairy farm products	0291
010200	Poultry and eggs	0251
010200	Poultry and eggs	0252
010200	Poultry and eggs	0253
010200	Poultry and eggs	0259
010200	Poultry and eggs	0191
010200	Poultry and eggs	0219
010200	Poultry and eggs	0291
010301	Meat animals	0211
010301	Meat animals	0212
010301	Meat animals	0213
010301	Meat animals	0214
010301	Meat animals	0219
010301	Meat animals	0191
010301	Meat animals	0259
010301	Meat animals	0291
010302	Miscellaneous livestock	0271
010302	Miscellaneous livestock	0272
010302	Miscellaneous livestock	0273
010302	Miscellaneous livestock	0279
010302	Miscellaneous livestock	0191
010302	Miscellaneous livestock	0219
010302	Miscellaneous livestock	0259
010302	Miscellaneous livestock	0291
020100	Cotton	0131
020100	Cotton	0191

primitive NAICS industries in the BLS employment matrices. We include a "1" in the (i,j)-th position in the concordance matrix if BEA or USAGE industry i in column 1 is concorded with NAICS industry j in column 3. Otherwise the (i,j)-th element in the concordance matrix is zero.

3. Step 3: allocating self-employed workers and un-paid family helpers to industries, 2004

The BLS data and forecasts include columns for 2004 and 2014 showing employment for self-employed workers and un-paid family helpers classified by occupation. In 2004 they numbered 12.1 million. For most occupations, we adjusted employment in the 324 primitive NAICS industries to take account of these workers according to the formula:

$$EN(o,k) = E_{step\,1}(o,k) * \left[1 + \frac{E_{SE}(o)}{\sum_h E_{step\,1}(o,h)} \right] \tag{3.1}$$

where

Table 2.3. A sample of the BEA/NAICS concordance

BEA	NAICS	NAICS in IndOcc04	NAICS Description
010100	11212	112000	Dairy Cattle and Milk Production
010200	11231	112000	Chicken Egg Production
010200	11232	112000	Broilers and Other Meat Type Chicken Production
010200	11233	112000	Turkey Production
010200	11239	112000	Other Poultry Production
010301	112111	112000	Beef Cattle Ranching and Farming (pt)
010301	112112	112000	Cattle Feedlots
010301	11221	112000	Hog and Pig Farming
010301	11241	112000	Sheep Farming
010301	11242	112000	Goat Farming
010302	112511	112000	Finfish Farming and Fish Hatcheries (pt)
010302	112512	112000	Shellfish Farming (pt)
010302	112519	112000	Other Animal Aquaculture (pt)
010302	112519	112000	Other Animal Aquaculture (pt)
010302	11291	112000	Apiculture
010302	11292	112000	Horses and Other Equine Production
010302	11293	112000	Fur-Bearing Animal and Rabbit Production
010302	11299	112000	All Other Animal Production (pt)
020100	11192	111000	Cotton Farming
020201	11114	111000	Wheat Farming
020201	11115	111000	Corn Farming (pt)
020201	11115	111000	Corn Farming (pt)
020201	11116	111000	Rice Farming
020201	111199	111000	All Other Grain Farming
020202	11115	111000	see above
020202	111199	111000	see above
020203	111199	111000	grass seeds
020300	11191	111000	Tobacco Farming
020401	11131	111000	Orange Groves

Table 2.4. Linking BEA with NAICS through SIC

BEA	SIC	NAICS
010100 Dairy farm products	0241 Dairy farms	112111 Beef cattle ranching and farming (pt) 11212 Dairy cattle and milk production
	0191 General farms, primarily crops	111998 All other misc. crop farming (pt)
	0259 Poultry and eggs, NEC	11239 Other poultry production
	0291 General farms, primarily livestock and animal specialties	11299 All other animal production (pt)

$E_{step\ 1}(o,k)$ is employment in occupation o, NAICS industry k estimated at the end of step 1;

EN(o,k) is employment in occupation o, NAICS industry k after adjustment for self employment and un-paid family workers; and

$E_{SE}(o)$ is the number of self-employed and un-paid family workers in occupation o.

For one occupation, Farmers and ranchers (o119012), formula (3.1) is clearly unsuitable. This occupation contains 1.1 million self-employed and un-paid family helpers but no wage and salary workers. Consequently (3.1) provides no basis for distributing non-wage-and-salary earners across industries in the same way as wage and salary earners. For farmers and ranchers we adopted the formula

$$EN(o119012,k) = E_{SE}(o119012) * \left[\frac{Labret\_N(k)}{\sum_{j \in Agr} Labret\_N(j)} \right] \quad , k \in Agr,$$

$$= 0 \quad , \quad k \notin Agr \tag{3.2}$$

where

Agr is the subset of agricultural industries in the NAICS 324 primitive industries (n111000 and n112000); and

Labret_N(k) is an estimate of the returns to labor in NAICS industry k, including imputed wages for self-employed and un-paid family workers. Labret_N(k) was estimated from USAGE data by aggregating labor returns for the USAGE agricultural industries that form NAICS industry k.

After the implementation of (3.1) and (3.2), we have a 753 by 324 matrix, EN, in which total employment is 145.3 million. This refers to civilian employment. Our matrix does not include people serving in the U.S. defense forces.

4. Step 4: preliminary estimate of the 753 by 513 employment matrix for 2004

In step 2 we developed a 513 by 324 concordance matrix of “1”s and “0”s. Now we create a matrix in which the “1”s are transformed into fractions and the “0”s are left unaltered. If component (u,n) in the original concordance matrix is a “1”, then we replace this component with a preliminary estimate [$F_1(u,n)$] of the share of USAGE industry u in employment of workers in NAICS industry n. Each column sum of the 513 by 324 F_1 matrix formed in this way is one.

For estimating the F_1 matrix we applied the formula

$$F_1(u,n) = \frac{Concord(u,n) * Labret\_U(u)}{\sum_q Concord(q,n) * Labret\_U(q)} \tag{4.1}$$

where

Concord(u,n) is the entry for USAGE industry u, NAICS industry n in the concordance matrix developed in step 2; and
Labret_U(u) is the return to labor in USAGE industry u in the 2004 USAGE database.

We use F_1 to make a preliminary estimate of employment by occupation and USAGE industry according to

$$EU_1(o,u) = \sum_n EN(o,n) * F_1(u,n) \quad . \qquad (4.2)$$

5. Step 5: using implied wage rates to refine the preliminary estimate of the 753 by 513 employment matrix for 2004

The assumption underlying (4.1) is that: if parts of USAGE industries s and r are in NAICS industry n, and returns to labor in s are twice those in r, then the share of employment in NAICS industry n that is accounted for by workers classified in USAGE industry s is twice that accounted for by workers classified in USAGE industry r. This assumption may not be satisfactory. Although s is a large employer relative to r, it may not be true that the part of s overlapping with n is large compared with the part of r overlapping with n.

As a check on the employment estimates obtained in (4.2) under assumption (4.1), we computed implied wage rates by USAGE industry as

$$W_1(u) = \frac{Labret\_U(u)}{\sum_o EU_1(o,u)} \quad . \qquad (5.1)$$

Many of the $W_1(u)$'s obtained this way seemed extreme, implying that we have allocated far too few or far too many workers to some USAGE industries. In response to this we reset the F matrix as:

$$F_2(u,n) = \frac{F_1(u,n) * W_1(u)}{\sum_q F_1(q,n) * W_1(q)} \quad . \qquad (5.2)$$

Where u is one of two or more USAGE industries that are partly included in NAICS industry n, then via (5.2), $F_2(u,n) > F_1(u,n)$ if $W_1(u)$ is high relative to the implied wage rates of other USAGE industries that are partly included in n. Thus, via (5.2) we assign more people to USAGE industry u if $W_1(u)$ is high, and less if $W_1(u)$ is low. Having reset the F matrix, we recomputed EU and W as

$$EU_2(o,u) = \sum_n EN(o,n) * F_2(u,n) \quad , \text{and} \qquad (5.3)$$

$$W_2(u) = \frac{Labret\_U(u)}{\sum_o EU_2(o,u)} \quad . \tag{5.4}$$

This recomputation produced much more realistic implied industry wage rates than those obtained in (5.1). We took this process to a third iteration, that is we computed F_3, EU_3 and W_3 as

$$F_3(u,n) = \frac{F_2(u,n) * W_2(u)}{\sum_q F_2(q,n) * W_2(q)} \quad . \tag{5.5}$$

$$EU_3(o,u) = \sum_n EN(o,n) * F_3(u,n) \quad \text{, and} \tag{5.6}$$

$$W_3(u) = \frac{Labret\_U(u)}{\sum_o EU_3(o,u)} \quad . \tag{5.7}$$

We found that there was very little movement from the estimates obtained in (5.2), (5.3) and (5.4). We accepted the third iteration results as the final outcome from the fifth step of our six-step estimating procedure.

6. Step 6: improving the 753 by 513 employment matrix for 2004 by giving the F matrix an occupation dimension.

On inspecting the EU_3 matrix generated at the end of step 5, we found some unsatisfactory results, especially for occupations for which government employment is important. For example, the EU_3 matrix indicated considerable employment of Education administrators, post secondary (occupation o119033) in all three USAGE State and local education industries (State and local consumption expenditure public school, SLCEpubSch; State and local consumption expenditure public higher education, SLCEpubHied; State and local consumption expenditure other education and libraries, SLCEothedLib). We judge that Education administrators, post secondary should be located predominantly in SLCEpubHied. So how did they find their way into the other two State and Local education industries?

The problem is that the NAICS industry classification used by the BLS has a single State and Local education industry (n946110) that encompasses the three USAGE industries. Thus in (5.6), F_3(u,n946110) > 0 for u equals SLCEpubSch, SLCEpubHied and SLCEothedLib and equals zero for all other u. While EU_3 gives SLCEpubSch, SLCEpubHied and SLCEothedLib different levels of aggregate employment reflecting their returns to labor in the USAGE database, it gives them identical occupational profiles. In particular, it implies that Education administrators, post secondary (o119033) accounts for an identical share of employment each of these three USAGE industries.

In responding to this problem, we started by drawing up a 753 by 19 concordance matrix, four rows of which are shown in Table 6.1.[5] The rows refer to occupations and the columns refer to the 19 government industries in the USAGE model. As indicated in the table, the 19 USAGE industries are encompassed by 4 NAICS industries. For each occupation we made a judgement[6] as to where its employment in NAICS government industries should be located among USAGE industries. These judgments are indicated by "1"s and "0"s in Table 6.1. For example, in the case of Education administrators, post secondary (o119033), we judged that all of State and Local education employment, that is employment allocated by the BLS to NAICS industry n946110, should be allocated to USAGE industry SLCEpubHied. This judgement is indicated by a "1" in column (6) of the o119033-row of Table 6.1 and zeros in columns (5) and (7). Secondly, with regard to o119033, we judged that all Federal government employment (that is employment in n919999) should be allocated to USAGE industry FCGEnondef. This judgement is indicated by a "1" in column (2) of the o119033-row and zeros in columns (1), (3) and (4). This second judgement is of relatively minor importance because few people in o119033 are employed in n919999. Almost no people in o119033 are employed in the final NAICS industry (n949400) shown in Table 6.1. Consequently we wasted no time in making a judgement about the distribution across USAGE industries of o119033 people employed in n949400. This lack of judgement is indicated by "1"s in all of the columns (9) to (19) of the o119033-row of Table 6.1. As we will see, these "1"s signify that o119033 people in n949400 should be distributed across all the USAGE components of n949400. No judgements are required with regard to the distribution to USAGE industries of NAICS employment in n946220. This is because we have assumed a one-to-one correspondence between n946220 and the USAGE industry State and local consumption expenditure, health (SLCEhealth).

---

[5] At the Centre of Policy Studies, the full version of Table 6.1 can be found in sheet 1 of our Excel workbook e:\usage\kyle\jan31\govtnew.xls.

[6] The time spent on these judgement varied across occupations reflecting the importance of the government as an employer.

Table 6.1. Selected rows from occupation/government-industry concordance matrix

	*	n919999				n946110			n946220	n949400										
occupation	Govt concentation	FGCEnatdef	FGCEnondef	FedElec	OthFedGovEnt	SLCEpubSch	SLCEpubHied	SLCEothedLib	SLCEhealth	SLCEwelfare	SLCEsanitat	SLCEpolice	SLCEfire	SLCEcorrect	SLCEhighway	SLCEnatural	Ptransit	Elec Util	OthS&Lent	SLCEother
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
o119033 Education administrators, post secondary	Y	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
o119111 Medical and health services managers	Y	1	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
o191023 Zoologists and wildlife biologists	Y	0	1	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0
o192043 Hydrologists	Y	1	1	0	0	1	1	1	1	0	1	0	0	0	1	1	0	0	0	0
o331011 First line supervisors/managers of correctional officers	Y	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0

* A "Y" in this column indicates that more than 20 per cent of the occupation's employment is in government industries. For occupations marked with a "Y", we paid particular attention to the allocation of their employment across USAGE government industries.

Next, we returned to the concordance matrix created in step 2. A “1” in (u,n)-th position of this matrix indicates that some of the people employed in NAICS industry n are employed in USAGE industry u. Now we add an occupation dimension. In the expanded matrix, a “1” in the (o,u,n)-th position indicates that some of the occupation-o people employed in NAICS industry n are employed in USAGE industry u. For all occupations o, we defined the entries in the expanded matrix, Concord_o, as

$$\text{Concord\_o}(o,u,n) = \begin{cases} \text{gov}(o,u) * \text{Concord}(u,n) & \text{if } u \in \text{GOV\_U and } n \in \text{GOV\_N} \\ \text{Concord}(u,n) & \text{otherwise} \quad , \end{cases} \quad \text{and} \qquad (6.1)$$

where

GOV_U and GOV_N are the 19 USAGE government industries and the 4 NAICS government industries identified in Table 6.1; and

gov(o,u) is the (o,u)-th component of the 753 by 19 matrix of “1”s and “0”s part of which is set out in Table 6.1.

To clarify (6.1) we work through an example. In the original concordance matrix,

$$\text{Concord(SLCEpubSch, n946110)} = 1 \quad , \qquad (6.2)$$

$$\text{Concord(SLCEpubHied, n946110)} = 1 \quad , \qquad (6.3)$$

$$\text{Concord(SLCEothedLib, n946110)} = 1 \quad , \qquad (6.4)$$

indicating that all 3 of the USAGE State and Local education industries are mapped to NAICS industry n946110. From Table 6.1 we see that NAICS industry n946110 is made up of USAGE industries SLCEpubSch, SLCEpubHied and SLCEothedLib and that

$$\text{gov(o119033, SLCEpubSch)} = 0 \quad , \qquad (6.5)$$

$$\text{gov(o119033, SLCEpubHied)} = 1 \quad , \qquad (6.6)$$

$$\text{gov(o119033, SLCEothedLib)} = 0 \quad , \qquad (6.7)$$

indicating that those Education administrators, post secondary who are employed in n946110 are located in SLCEpubHied, but not in SLCEpubSch or SLCEothedLib. Now applying (6.1) we find that

$$\text{Concord\_o(o119033, SLCEpubSch, n946110)} = 0 \quad , \qquad (6.8)$$

$$\text{Concord\_o(o119033, SLCEpubHied, n946110)} = 1 \quad , \qquad (6.9)$$

$$\text{Concord\_o(o119033, SLCEothedLib, n946110)} = 0 \quad . \qquad (6.10)$$

Equation (6.9) indicates that some of the occupation-o119033 people employed in NAICS industry n946110 are employed in USAGE industry SLCEpubHied. Equations (6.8) and (6.10) indicate that none of the occupation-o119033 people employed in NAICS industry n946110 is employed in USAGE industries SLCEpubSch and SLCEothedLib. Together, (6.8) to (6.10) imply that all of the occupation-o119033 people employed in NAICS industry n946110 are employed in USAGE industry SLCEpubHied.

Now we use the expanded concordance matrix to make a preliminary estimate of the share, $F0_1(o,u,n)$, of USAGE industry u in employment of occupation-o workers from NAICS industry n:

$$FO_1(o,u,n) = \frac{Concord\_o(o,u,n) * Labret\_U(u)}{\sum_q Concord\_o(o,q,n) * Labret\_U(q)} \quad . \tag{6.11}$$

The assumption underlying (6.11) is that: if some occupation-o workers in NAICS industry n are located in USAGE industries s and r, and returns to labor in s are twice those in r, then the share of o-worker employment in n that is accounted for by workers classified in USAGE industry s is twice that accounted for by o-workers classified in USAGE industry r.

From here, we proceed as in steps 4 and 5. We compute

$$EOU_1(o,u) = \sum_n EN(o,n) * FO_1(o,u,n) \quad . \tag{6.12}$$

where

$EOU_1(o,u)$ is an estimate of employment in occupation o and USAGE industry u incorporating our occupation-expanded concordance matrix.

EOU_1 is an improvement on our earlier matrix EU_1 because it reflects the judgements built into Table 6.1. For example, it does not misallocate Education administrators, post secondary to USAGE industries SLCEpubSch and SLCEothedLib.

As in step 5, we computed implied wage rates

$$WO_1(u) = \frac{Labret\_U(u)}{\sum_o EOU_1(o,u)} \quad , \tag{6.13}$$

we re-set FO,

$$FO_2(o,u,n) = \frac{FO_1(o,u,n) * W_1(u)}{\sum_q FO_1(o,q,n) * W_1(q)} \quad , \tag{6.14}$$

and we re-estimated EOU,

$$EOU_2(o,u) = \sum_n EN(o,n) * FO_2(o,u,n) \quad . \tag{6.15}$$

Again, as in step 5 we proceeded through a third iteration, obtaining EOU_3. This is close to our final estimate of employment by occupation and USAGE industry.

As we did with EU_3, we inspected EOU_3 looking for implausibilities. An example that we noticed was a considerable allocation of Farmers and ranchers (o119012) to the USAGE industry Agricultural, forestry and fishery services (u040001). This occurred because part of u040001 is contained in the NAICS industry Animal production (n112000), which is a major employer of Farmers and ranchers. To ensure that Farmers and ranchers are not allocated to u040001, we repeated our computations starting with a modified concord_o matrix in which

$$Concord\_o(o119012,u040001,n112000) = 0 \quad . \tag{6.16}$$

Undoubtedly further implausibilities will be noticed when our occupation by USAGE-industry matrices are applied in USAGE simulations. Using the GEMPACK programs created for the manipulations described in this paper, it will not be difficult to eliminate such implausibilities.

As in our computations of EU_i , i = 1, …, 3 in section 5, we found in our computations of EOU_i , i = 1, …, 3 that implied industry wage rates were extreme for EOU_1 but seemed reasonable for EOU_2. Out of the 508 USAGE industries that employ labor, the implied wage rates in EOU_2 for 429 were between half and twice the economy-wide average implied wage rate. EOU_3 is very similar to EOU_2. In EOU_3, 432 of the implied wage rates were between half and twice the economy-wide average implied wage rate.

7. USAGE employment matrix for 2014

We derived a 753 by 513 employment matrix for 2014 by applying the six steps outlined above to the BLS employment matrix for 2014. Steps 3 to 6 require wagebill data by USAGE industry for 2014 to form various shares. We developed the required wagebills for 2014 by extrapolating forecasts of movements in wagebills for 2004 to 2010 out to 2014.

8. Comparing occupation/industry employment matrices for 2004 and 2014

We suspect that BLS forecasts for employment by occupation and industry are formed largely as sums of industry effects and occupation effects. To check this hypothesis and to facilitate comparison between our 753 by 513 employment matrices for 2004 and 2014, we ran the regression

$$g(o,j) = \alpha(o) + \beta(j) + u(o,j) \quad , \tag{8.1}$$

where

g(o,j) is growth between 2004 and 2014 in employment in occupation o and industry j;
$\alpha(o)$ and $\beta(j)$ are parameters; and
u(o,j) is the disturbance term.

That is, we tried to explain growth in employment in occupation o in industry j as the sum of an occupation effect and an industry effect.

Some of the growth rates for occupation/industry cells containing very few people (less than 10 in both 2004 and 2014) are extremely large in absolute terms. This is of no practical importance but it prevents the simple equation (8.1) from revealing anything of interest. To deal with the problem of small cells we formed a weighted version of (8.1):

$$z(o, j) = \alpha(o) * E(o, j)^{0.5} + \beta(j) * E(o, j)^{0.5} + u(o, j) \quad , \tag{8.2}$$

where

E(o,j) is employment in occupation o and industry j in 2004;
z(o,j) equals $g(o,j)*E(o,j)^{0.5}$; and
u(o,j) is the disturbance term.

This regression fits with an R-squared of 0.93.

In interpreting the values for the $\alpha(o)$'s and $\beta(j)$'s it is necessary to adopt a normalisation convention. Clearly, if $\hat{\alpha}(o)$ and $\hat{\beta}(j)$ for all o and j minimize the sum of the squared residuals in (8.2), then $[\hat{\alpha}(o) + \lambda]$ and $[\hat{\beta}(j) - \lambda]$ also minimize the sum of squared residuals for any value of λ. We adopted the normalization that the average value of the α's is zero, that is

$$\sum_{o} \alpha(o) * \frac{E(o,\bullet)}{E(\bullet,\bullet)} = 0 \quad , \tag{8.3}$$

where

$E(o,\bullet)$ is employment in occupation o in 2004; and
$E(\bullet,\bullet)$ is aggregate employment in 2004.

Under this normalization, for most industries $\hat{\beta}(j)$ is close to the growth in employment in industry j. Thus $\hat{\alpha}(o)$ is the growth premium for occupation o. For example, we can interpret $\hat{\alpha}(o) = 2$ as indicating that growth in employment for occupation o in every industry between 2004 and 2014 exceeds the industry growth rate by about 2 percentage points.

Table 8.1. Selected results for occupation growth premiums, $\hat{\alpha}(o)$

No.	Occupation	Model (8.2)	Model (8.4)
o151081	Network systems and data communications analysts	39.88	38.25
o151031	Computer software engineers, applications	30.13	26.91
o151032	Computer software engineers, systems software	28.18	25.00
o194092	Forensic science technicians	26.02	22.56
o151071	Network and computer systems administrators	23.98	22.44
o151061	Database administrators	23.52	22.08
o291071	Physician assistants	23.49	21.85
o319092	Medical assistants	21.90	20.66
o311011	Home health aides	20.59	14.72
o172031	Biomedical engineers	20.24	17.71
o439071	Office machine operators, except computer	-33.29	-32.26
o432021	Telephone operators	-36.14	-37.75
o534099	Rail transportation workers, all other	-39.38	-36.62
o534021	Railroad brake, signal, and switch operators	-45.48	-42.35
o439011	Computer operators	-47.12	-47.51
o435041	Meter readers, utilities	-49.07	-49.24
o519132	Photographic processing machine operators	-49.41	-42.60
o434041	Credit authorizers, checkers, and clerks	-50.53	-49.84
o439051	Mail clerks and mail machine operators, except postal service	-53.82	-52.75
o434071	File clerks	-57.64	-57.56

Table 8.1 contains the ten highest and lowest $\hat{\alpha}(o)$'s estimated from the model in equation (8.2). The group of occupations enjoying large positive growth premiums is dominated by computer professionals. It also includes four medical occupations. At the other extreme, the group of occupations suffering large negative growth premiums is dominated by clerical workers. The losing group also includes computer operators, reflecting the rapid movement towards personal computers operated by non-specialists.

In addition to (8.2), we fitted the equation

$$d(o, j) = \alpha(o) * E(o, j)^{0.5} + u(o, j) \quad , \qquad (8.4)$$

where

E(o,j) is employment in occupation o and industry j in 2004;
d(o,j) equals $[g(o, j) - g(\bullet, j)] * E(o, j)^{0.5}$ with $g(\bullet, j)$ being the growth rate in total employment in industry j; and
u(o,j) is the disturbance term.

Via (8.4) we again check the hypothesis that the BLS forecasts for the growth in the share of occupation o in an industry's employment is largely independent of the industry. The R-squared for (8.4) was 0.67 and, as can be seen from Table 8.1, the results for the $\hat{\alpha}$'s are quite similar to those obtained in model (8.2).

9. Employment by occupation and industry in the USAGE model: estimating occupational-saving technical progress (April 17, 2006 updated May 3, 2006)

We assume in USAGE that industry j chooses E(r,j) to minimize

$$\sum_{r} E(r, j) * W(r) \quad , \qquad (9.1)$$

subject to

$$E(j) = CES_j\left(\frac{E(r, j)}{A(r, j)}, r \in OCC\right) \quad , \qquad (9.2)$$

where

E(r,j) is employment in occupation r and industry j;
E(j) is a measure of overall labour input in industry j;
W(r) is the wage rate for occupation r;
OCC is the set of all occupations; and
A(r,j) is a technology variable. A 10 per cent reduction in A(r,j) means that industry j can achieve a given level of overall labor input with 10 per cent less occupation r hours and no change in the input of any other occupation.

Model (9.1)-(9.2) generates labour demand equations that can be included in USAGE in percentage change form as:

$$e(r, j) = e(j) - \sigma_j\left(w(r) - \sum_{o} S(o, j) * w(o)\right) + a(r, j) - \sigma_j\left(a(r, j) - \sum_{o} S(o, j) * a(o, j)\right) \quad , \qquad (9.3)$$

where

lowercase symbols, e(r,j), e(j), w(r) and a(r,j), are percentage changes in the variables represented by corresponding uppercase symbols;
S(o,j) is the share of j's labor costs accounted for by workers in occupation o; and
σ_j is the elasticity of substitution between different occupations in industry j.

We define the average occupation-saving technical change in industry j as

$$a\_ave(j) = \sum_{o} S(o, j) * a(o, j) \quad . \qquad (9.4)$$

In USAGE we assume that a_ave(j) is zero for all j. There is no loss of generality in this assumption because we allow for overall labour-saving technical change in the determination of e(j). We restrict the role of a(r,j)s to simulating the effects of biases between occupations in labor-saving technical change.

From the BLS we obtain wage rates [W(r)] by occupation for 2002 from Table I-1 at http://www.bls.gov/emp/optd/home.htm (stored as OPTD_MTB.xls on RHS computer). On the assumption that there was no change in occupational wage relativities between 2002 and 2004, we used the BLS occupational wage rates, together with our 753 by 513 employment matrix for 2004 to compute S(o,j)s.

To estimate percentage movements in the A(r,j)s for the period 2004 to 2014, we used model (9.3) and (9.4) with:

σ_j set at 0.35 for all j (the number adopted in MONASH and at least temporarily in USAGE);
a_ave(j) treated exogenously and set at zero;
w(r) treated exogenously and set at zero (this is equivalent to assuming that the BLS is anticipating no change in occupational wage relativities);
e(r,j) treated exogenously and set at the percentage movement forecast for 2004 to 2014; and
a(r,j) treated endogenously.

Via (9.3) and (9.4) we obtained estimates for the a(r,j)s that are effectively free of linearization error by using a 16-step Euler computation.

To describe our results for the a(r,j)s we fitted a regression similar to (8.4)[7]:

$$b(o,j) = \alpha(o) * E(o,j)^{0.5} + u(o,j) \quad , \qquad (9.5)$$

where

E(o,j) is employment in occupation o and industry j in 2004;
b(o,j) equals $a(o,j) * E(o,j)^{0.5}$; and
u(o,j) is the disturbance term.

Via (9.5) we yet again check the hypothesis that the BLS forecasts for the growth in the share of occupation o in an industry's employment is largely independent of the industry. The R-squared for (9.5) was 0.68.

Table 9.1 lists the ten occupations with the highest estimated $\alpha(o)$s from (9.5) and the lowest estimated $\alpha(o)$s. There is a complete overlap between the bottom-ten occupations identified in Table 9.1 and the bottom-ten identified in Table 8.1. The overlap at the top ends of Tables 9.1 and 8.1 is less pronounced than that at the bottom end. Three occupations [Hunters and trappers; Cutters and trimmers, hand; and Fashion designers] appear in the top ten in Table 9.1 that were not in the top ten in Table 8.1. However, none of these occupations missed out by much from being in the top 10 in Table 8.1. Similarly, three occupations [Medical assistants; Home health aides; and Biomedical engineers] appear in the top ten in Table 8.1 that are not in the top ten in Table 9.1. These occupation just missed out from being in the top ten in Table 9.1.

---

[7] The computation of this regression is stored at E:\usage\kyle\jan31\growth3.tab.

Table 9.1. Selected results for occupation technology shifts $[\hat{\alpha}(o)]$, 2004 to 2014

No.	Occupation	Model (9.5)
o453021	Hunters and trappers	57.86
o151081	Network systems and data communications analysts	50.95
o151031	Computer software engineers, applications	31.83
o194092	Forensic science technicians	30.50
o151032	Computer software engineers, systems software	30.11
o519031	Cutters and trimmers, hand	28.88
o291071	Physician assistants	28.66
o151071	Network and computer systems administrators	28.10
o271022	Fashion designers	27.99
o151061	Database administrators	27.68
o439071	Office machine operators, except computer	-42.41
o534099	Rail transportation workers, all other	-48.63
o432021	Telephone operators	-52.25
o519132	Photographic processing machine operators	-53.85
o534021	Railroad brake, signal, and switch operators	-55.60
o439011	Computer operators	-57.79
o439051	Mail clerks and mail machine operators, except postal service	-62.04
o434041	Credit authorizers, checkers, and clerks	-62.61
o435041	Meter readers, utilities	-63.61
o434071	File clerks	-64.30

Why does Hunters and trappers have such a large α value in Table 9.1? In the BLS forecasts, Hunters and trappers (o453021) are employed entirely in Fishing, hunting and trapping (NAICS industry n114000). For the period 2004 to 2014 the BLS project 4.08 per cent growth in employment for Hunters and trappers and a 20.57 per cent decline in employment in Fishing, hunting and trapping. Thus, it was inevitable that we would obtain for α(o453021) a value of about 52 per cent [=100*{(1.0408/(1-0.2057))$^{(1/0.65}$ - 1}]. In fact we get a higher value, 57.86. The USAGE industry that employs Hunters and trappers is Forestry products (b30001). After mapping the NAICS data into the USAGE classifications we find that the BLS forecasts imply a decline in employment in b30001 measured in number of jobs of 14.3 per cent. However in wagebill terms the decline in labor input is 23.14 per cent, that is e(b30001) in (9.3) is -23.14. Thus, α(o453021) turns out to be about 59.4 per cent [=100*{(1.0408/(1-0.2314))$^{(1/0.65}$ - 1}].[8]

The discrepancy between the movement in jobs (-14.3 per cent) and the movement in labor input (-23.14 per cent) in b30001 explains how Hunters and trappers can appear at the top of Table 9.1 while not appearing in Table 8.1. The α's in Table 8.1 reflect the growth in employment of Hunters and trappers relative to growth in jobs in b30001. The α's in Table 9.1 reflect the growth in employment of Hunters and trappers relative to growth in labor input in b30001.

In applying USAGE to project demands for labor by occupation, our plan is to use the estimated value of $\alpha(o)$, prorated for the relevant number of years, as the exogenously set value for a(o,j) in (9.3). For example, in simulations for 2004 to 2010 we will have

$$a_{04-10}(o, j) = 100 * [(1 + \hat{\alpha}(o) / 100)^{6/10} - 1] \quad , \; o \in OCC, j \in IND . \qquad (9.6)$$

In using (9.6), we assume that the occupation biases deduced from the ten-year BLS forecasts can be prorated to six-year forecasts. We also assume that occupation-bias effects are independent of industries [there is no j on the RHS of (9.6)]. Some of the a(o,j)s calculated from (9.3) and (9.4), especially for o,j cells in which employment is very small, are extreme. By using average values [the $\hat{\alpha}(o)$s] we avoid introducing these extreme and unreliable estimates into our projections.

10. A comparison of BLS and USAGE forecasts for 2004 to 2010 (May 6, 2006)

We generate USAGE forecasts of growth by occupation, $g_{0410}^{USAGE}(o)$, for 2004 to 2010 using (9.3) with the a(o,j)s set according to (9.6). We also compute implied BLS forecasts, $g_{0410}^{BLS}(o)$, for this period as:

$$g_{0410}^{BLS}(o) = 100 * \left(\left(\frac{E_{14}(o)}{E_{04}(o)} \right)^{0.6} - 1 \right) \quad , o= 1, \ldots, 753,$$

where $E_{04}(o)$ and $E_{14}(o)$ are employment in occupation o in 2004 and 2014 in the BLS data and forecasts. Figures 10.1 and 10.2 are scatter diagrams of the two sets of forecasts. Figure 10.1

[8] The gap between 59.4 and 57.86 is probably explained by linearization errors in our estimation of the a(o,i)s.

includes all 753 occupations. In Figure 10.2 we have omitted the 26 occupations for which the points in Figure 10.1 are furthest from the regression line.

Figures 10.1 and 10.2 show that the USAGE and BLS forecasts are quite well correlated, with a few outstanding outliers. The differences between the two sets of occupational forecasts are caused almost entirely by differences between BLS and USAGE forecasts for labour input by industry. This is established in Figure 10.3. There we show that there is almost no difference between the BLS forecasts and USAGE forecasts generated under (9.3) and (9.6) but with e(j) in (9.3) set according to the BLS forecasts of industry labor input.[9]

Figure 10.4 compares USAGE and BLS forecasts of labor input by industry. To understand the differences between the BLS and USAGE occupational forecasts all that is necessary to understand is the differences between the BLS and USAGE industry forecasts illustrated in this figure.

In future research we intend to study Figure 10.4 in detail. However, in the time that we currently have available we will merely comment on two outliers.

Commercial fishing

BLS sees quite weak prospects, even weaker than recent history. Do they have some information on environmental or other restrictions? We see strong prosects. Commercial fishing is heavily trade exposed: 45 per cent of output is exported and imports represent about 75 per cent of domestic sales. We expect commercial fishing to be strongly favoured by devaluation.

Home health

BLS sees strong prospects, stronger than recent history. They argue that Home health services are becoming increasingly popular as the population ages. We see weaker prospects than recent history. In our model the slowdown in employment in home health is caused by: a reduction in the rate of growth in aggregate consumption; and an increase in the rate of growth in the price of Home health.

---

[9] The percentage change in labor input in industry j is a weighted average of the percentage changes in employment by occupation with the weights being occupational shares in industry labor costs. The BLS does not forecast labor input directly. We computed labor inputs for the 513 USAGE industries from the 753 by 513 employment matrices derived earlier in this paper using the S(o,j)s described in section 9.

Figure 10.1. Percentage growth in employment by occupation, 2004-10: comparison of BLS and USAGE forecasts

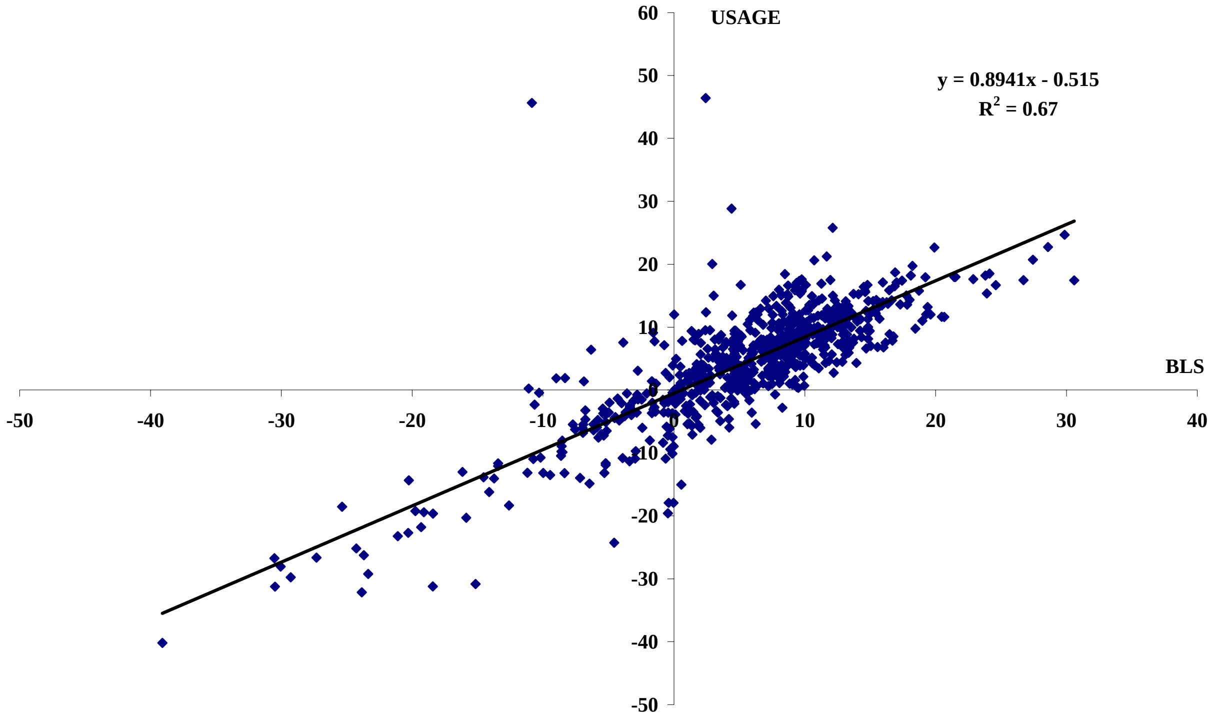

Figure 10.2. Percentage growth in employment by occupation, 2004-10: comparison of BLS and USAGE forecasts without outliers

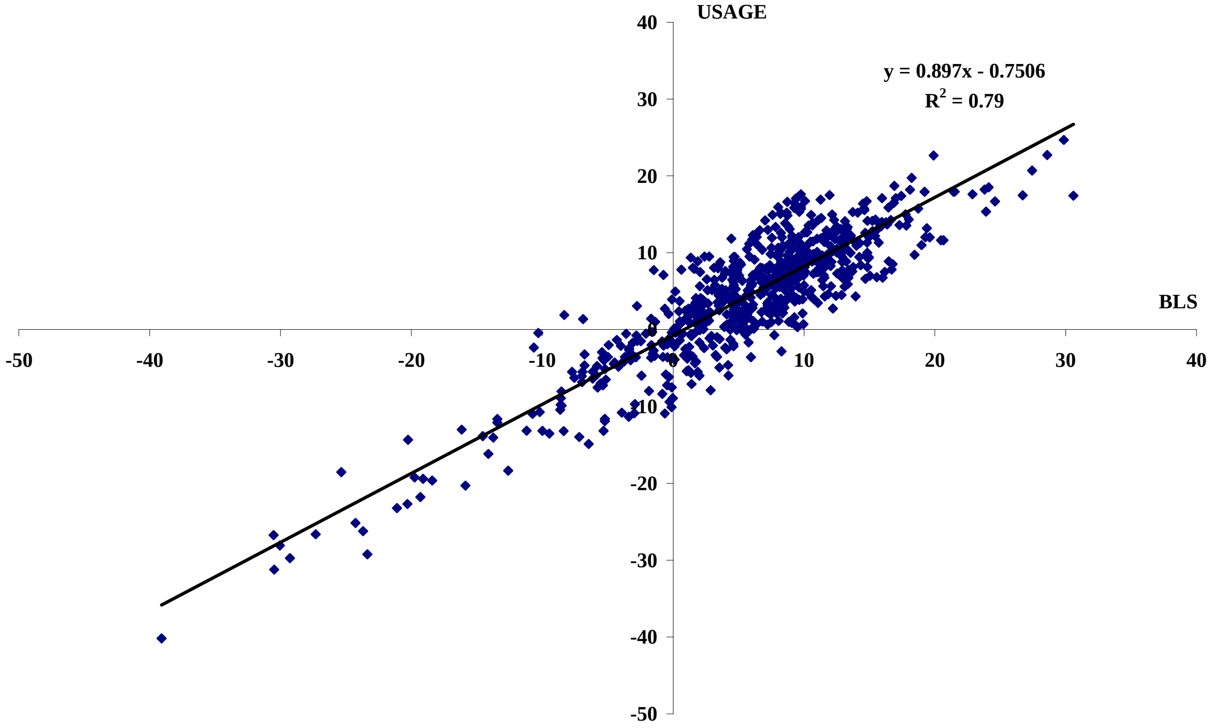

Figure 10.3. Percentage growth in employment by occupation, 2004-10: comparison of BLS occupation forecasts with USAGE forecasts incorporating BLS industry forecasts

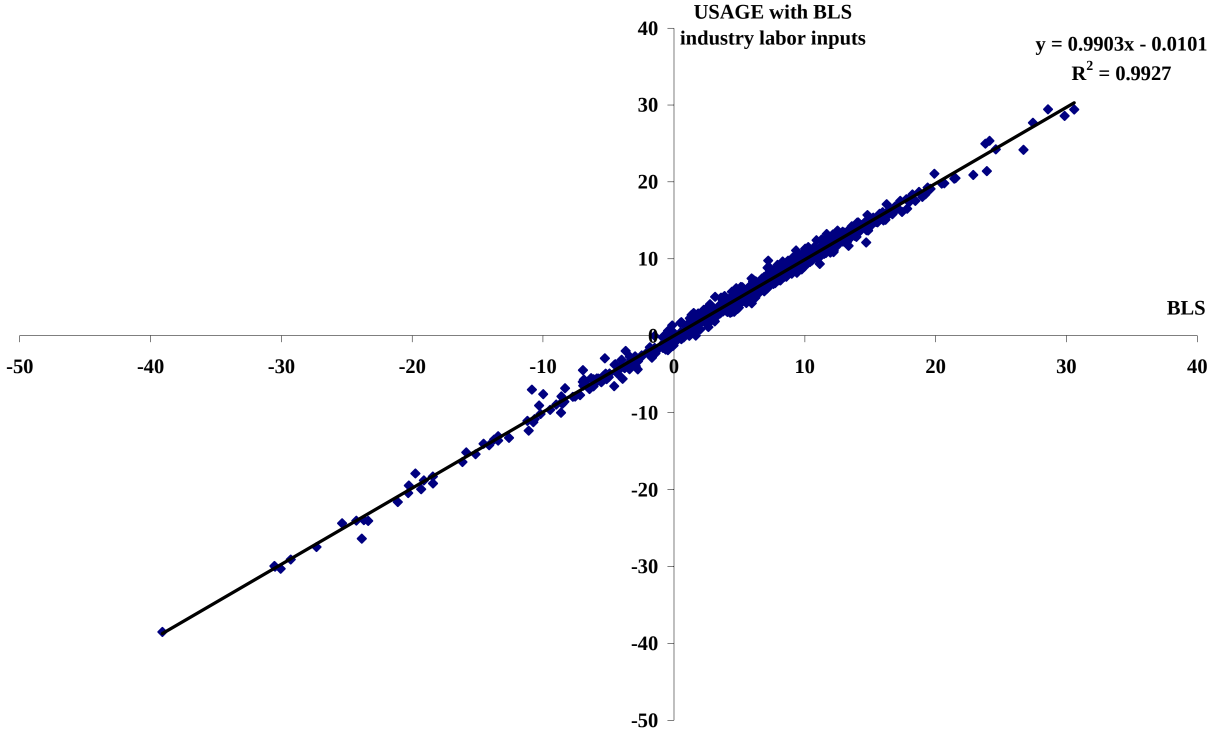

Figure 10.4. Percentage growth in labor input by industry, 2004-10: comparison of BLS forecasts with USAGE forecasts