**DISAGGREGATE DYNAMICS AND ECONOMIC GROWTH IN
CANADA**

by

ELIZABETH CLARE WAKERLY

B.A., University of Sheffield, 1987

M.A., Carleton University, 1988

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Economics

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

January 1997

©Elizabeth Clare Wakerly, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of the thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of the thesis for financial gain shall not be allowed without my written permission.

Department of Economics
The University of British Columbia
997-1873 East Mall
Vancouver, B.C., Canada
V6T 1Y2

Date:

10 Nov 1997

Abstract

This thesis takes the form of three essays in which I use disaggregate and aggregate information to examine Canadian economic growth.

In the first essay, I present evidence that the process of economic growth differs for low income per capita provinces and industries. This contrasts with results from traditional studies of economic convergence. In those papers, estimates of a rate of convergence suggest that poor provinces eventually “catch up” to rich provinces by growing faster. Unfortunately, this approach ignores the pattern of economic growth within the cross-section distribution. Explicitly modelling the evolving distribution, I find little mobility in the cross-sectional ordering and some evidence of divergence. In the long run, the poor stay (relatively) poor and the rich remain (relatively) rich.

In the second essay, I examine the dynamic effects of aggregate and disaggregate disturbances on both economic growth and the interaction between disaggregates. The approach is motivated by the class of models which predict two-way interaction between aggregate and disaggregate behaviour, such as Durlauf [28]. The disaggregate disturbance is identified as having no long-run impact on aggregate economic growth. I find that the aggregate shock has a large impact on aggregate income in both the short and long run; and accounts for most of its variation. The disaggregate shock contains some information for aggregate activity at business cycle horizons. Most interaction is explained by the disaggregate disturbance; the aggregate shock contributes little.

In the third essay, I present results from a variety of unit root tests on provincial and manufacturing industry panel income data. Standard Dickey-Fuller unit root tests applied to panels require averaging of data across the cross-section. More powerful tests allow pooling of cross-section and time-series information. Using these methods, I find that the null hypothesis of a unit root is rejected—strongly contrasting with results obtained using the standard Dickey-Fuller methodology.

Contents

Abstract	ii
Contents	iii
List of Tables	v
List of Figures	vii
Acknowledgement	ix
Chapter 1 Introduction	1
Chapter 2 The Evolving Distribution of Canadian Economic Growth	6
2.1 Introduction	6
2.2 Related Literature	8
2.3 Econometric Model	14
2.4 Cross-Section Dynamics	15
2.4.1 Provincial	16
Personal income	16
Wages, salaries and supplementary labour income	19
2.4.2 Industrial	21
2.5 Robustness	24
2.5.1 Provincial data adjustment	25
Relative to Ontario (1)	25
Relative to Ontario (2)	26
Aggregating some provinces	26
2.5.2 Industry data	27
Ontario	27
Western provinces	28
Eastern provinces	28
2.6 Conclusions	29
Chapter 3 The Dynamic Effects of Aggregate and Disaggregate Dis-	
turbances	37
3.1 Introduction	37
3.2 Identification	39
3.3 Interpretation	40
3.4 Results	41
3.4.1 Provincial data	41
3.4.2 Industrial data	45
3.5 Conclusions	46
Chapter 4 Testing For Canadian Unit Roots: A Panel Data Approach	57
4.1 Introduction	57

4.2 Unit Roots and Panel Data	58
4.3 Methods	62
4.3.1 The panel Dickey-Fuller test	62
4.3.2 The δ -bar test	63
4.3.3 The t -bar test	67
4.4 Results	69
4.4.1 Provincial data	69
Panel Dickey-Fuller test	69
δ -bar test: base case	70
δ -bar test: index adjustment	71
t -bar test	71
4.4.2 Industrial data	73
Panel Dickey-Fuller test	73
δ -bar test: base case	73
δ -bar test: index adjustment	73
t -bar test	74
4.4.3 Dickey-Fuller test	76
4.5 Conclusions	77
Chapter 5 Conclusions	79
Bibliography	83
Appendices	89
A The Data	90
A.1 Data Sources	90
A.2 Manufacturing Industries	93
B Fractile Transition Probability Matrices	94
C Further Estimations	95
C.1 Provincial Economic Growth per Capita	95
C.2 Provincial WSSLI per Capita	96
C.3 Provincial Employment per Capita	97
C.4 Industrial GDP per Employee	98
D Tests For Robustness	107
E Results Using Coulombe and Lee Data	117
E.1 Data	117
E.2 Mobility Dyanmics	118
E.2.1 Output and income per capita	118
E.2.2 Productivity measures	119
E.2.3 Longer sample period	120
E.3 Conclusions	120
F Quah and Sargent's Dynamic Index Model	122
F.1 Theoretical Model	122
F.2 Application	124

List of Tables

2.1	Transition Probability Matrix—Personal Income per Capita by Province	16
2.2	Fractile Transition Probability Matrix—Personal Income per Capita by Province	18
2.3	Transition Probability Matrix—Wages, Salaries and Supplementary Labour Income per Employee, by Province	20
2.4	Transition Probability Matrix—GDP per Employee, 21 Manufacturing Industries	23
3.1	Variance Decomposition of Aggregate Income per Capita and Provincial Interaction	44
3.2	Variance Decomposition of Aggregate GDP per Employee and Manufacturing Industry Interaction	47
4.1	δ -Bar Test Statistics for Provincial Data	70
4.2	δ -Bar Test Statistics (with Index Adjustment) for Provincial Data	71
4.3	t -Statistics (t -Bar Test) for Provincial Data	72
4.4	δ -Bar Test Statistics for Manufacturing Industry Data	74
4.5	δ -bar Test Statistics (with Index Adjustment) for Manufacturing Industry Data	74
4.6	t -Statistics (t -Bar Test) for Manufacturing Industry Data	75
4.7	DF Test Statistics for Provincial and Manufacturing Industry Data	76
B1	WSSLI Per Employee By Province	94
B2	GDP Per Employee, 21 Manufacturing Industries	94
C1	Transition Probability Matrix—Income Growth by Province	95
C2	Transition Probability Matrix—WSSLI Per Capita by Province	97
C3	Transition Probability Matrix—Employment Per Capita by Province	98
C4	Transition Probability Matrix—GDP Per Person at Work by Industry Sector	99
C5	Transition Probability Matrix—GDP Per Person Hour at Work by Industry Sector	99
D1	Transition Probability Matrix—Income by Province, Relative to Ontario	107
D2	Transition Probability Matrix—Income by Province, Subtracting Ontario	108
D3	Transition Probability Matrix—Income by Province, Aggregating some Provinces	108

D4 Transition Probability Matrix—GDP Per Employee, Manufacturing Industries, Ontario	109
D5 Transition Probability Matrix—GDP Per Employee, Manufacturing Industries, Western Provinces	109
D6 Transition Probability Matrix—GDP Per Employee, Manufacturing Industries, Eastern Provinces	110
E1 Data Used in Coulombe and Lee [21]	117
E2 Data Used in Coulombe and Lee [47]	118
E3 Data Used in Coulombe and Lee [22]	118

List of Figures

Figure 1 Distribution Dynamics	31
Figure 2 Log of Provincial Personal Income per Capita (1926-1994) . .	32
Figure 3 Quantiles for Log of Provincial Personal Income per Capita (1928-1992)	33
Figure 4 Log of Provincial WSSLI per Employee (1966-1994)	34
Figure 5 Quantiles for Log of Provincial WSSLI per Employee (1968- 1993)	35
Figure 6 Log of Manufacturing Industry GDP per Employee (1961:1- 1994:11)	36
Figure 7 Impulse Response: Aggregate Disturbance on Aggregate In- come Per Capita	49
Figure 8 Impulse Response: Disaggregate Disturbance on Aggregate Income per Capita	50
Figure 9 Impulse Response: Aggregate Disturbance on Interaction . .	51
Figure 10 Impulse Response: Disaggregate Disturbance on Interaction	52
Figure 11 Impulse Response: Aggregate Disturbance on Aggregate GDP per Employee	53
Figure 12 Impulse Response: Disaggregate Disturbance on Aggregate GDP per Employee	54
Figure 13 Impulse Response: Aggregate Disturbance on Interaction Measure	55
Figure 14 Impulse Response: Disaggregate Disturbance on Interaction Measure	56
Figure C1 Growth of Provincial Personal Income per Capita (1927-1993)	100
Figure C2 Quantiles for Growth of Provincial Personal Income per Capita (1929-1992)	101
Figure C3 Log of Provincial WSSLI per Capita (1961-1992)	102
Figure C4 Quantiles for Log of Provincial WSSLI per Capita (1963-1991)	103
Figure C5 Provincial Employment per Capita (1966-1993)	104
Figure C6 Index of GDP per Employee 1986=100 (1961-1994)	105
Figure C7 Index of GDP per Person Hour Worked 1986=100 (1961-1994)	106
Figure D1 Quantiles for Log of Provincial Personal Income per Capita, Divided by Ontario Data (1928-1993)	111
Figure D2 Quantiles for Log of Provincial Personal Income per Capita, Subtracting Ontario Data (1928-1993)	112

Figure D3 Quantiles for Log of Aggregated Provincial Personal Income per Capita (1928-1991)	113
Figure D4 Quantiles for Log of Manufacturing Industry GDP per Employee: Ontario (1973-1990)	114
Figure D5 Quantiles for Log of Manufacturing Industry GDP per Employee: Western Provinces (1973-1990)	115
Figure D6 Quantiles for Log of Manufacturing Industry GDP per Employee: Eastern Provinces (1973-1990)	116
Figure F1 Sample Standard Deviations: Including Employment (1968-1988)	125
Figure F2 Sample Standard Deviations: Including GNP (1968-1988)	126
Figure F3 Sample Standard Deviations: Including Employment vs including GNP (1968-1988)	127
Figure F4 Sample Standard Deviations: Two-Index Model vs Total Employment (1968-1988)	128
Figure F5 Sample Standard Deviations: Two-Index Model vs Total GNP (1968-1988)	129

Acknowledgement

I would like to thank my thesis committee, John Cragg, John Helliwell and James Nason for helpful comments and guidance. I am indebted to my fellow graduate students and my parents for their support. In particular, I would like to thank Shaun Vahey, without whose support, encouragement and patience this research would not have been possible.

Chapter 1

Introduction

This thesis takes the form of three essays in which I use aggregate and disaggregate information to examine Canadian economic growth. The importance of disaggregate information in explaining aggregate activity is emphasised in the papers by Durlauf [28], Galor and Zeira [35], Long and Plosser [53] and Quah [68]. I find that, in general, averaging out disaggregate effects causes a loss of information and generates misleading results.

In the first essay, Chapter 2, I examine the contribution of disaggregates to economic convergence. Traditionally, empirical work focuses on estimating a single rate of convergence across disaggregates. Using Canadian provincial data and techniques Barro [3] and Barro and Sala-i-Martin [4] develop, various authors, that include Coulombe and Lee [21] and [47], Helliwell [40] and Sala-i-Martin [77], regress growth rates on initial income per capita levels. They find convergence of around 1-3% per annum for selected post-WWI samples. An interpretation of this result takes the form of evidence that, in the long run, poor regions "catch up" rich regions (or the differential between the rich and poor decreases) by growing faster.

Unfortunately, the approach used in these studies reveals little about the evolving pattern of economic growth within the cross-section because averaging annihilates the disaggregate information. A convergence rate of say 2% per annum across Canadian provinces, sheds no light on whether rich and poor economies possess different short-run growth dynamics. Nor does it provide any information about the probability of a relatively poor province improving its rank in the cross-section. (And arguably, both policy makers and individuals are more concerned with rankings than with absolute economic growth.) An average convergence rate cannot tell us whether it is the already rich which are converging towards each other or whether the poor are becoming relatively poorer.

Existing studies also neglect the relationship between provincial and industrial patterns of growth. If, for example, the pattern of economic growth for the major manufacturing industries in Newfoundland and Alberta—food and chemicals respectively—is very different, then the finding that income per capita in the two provinces is converging should be treated with some caution. Since the growth patterns of the leading industries differ, the path of provincial income per capita most likely differs as well.

As an alternative to the traditional approach to measuring convergence, I explicitly study the relationships among Canadian disaggregates. To do this, I model the cross-section dynamics of Canadian income at a disaggregate level. The approach I follow builds on work by Quah [69]. I use disaggregate provincial and industrial data, and examine the sensitivity of the results to different measures of economic growth, to the inclusion of Ontario data, and to asymmetries in province size.

I show that although the long-run cross-section distribution of provincial income per capita is generally unimodal, there exists virtually no mobility in the cross-sectional ordering of income levels. I also report evidence to support the view that the process of economic growth is distinct for poorer regions. Put differently, poor provinces remain relatively poor and rich provinces remain relatively rich. Adjusting the data to account for the size differences of the provinces, and for the disproportionate role played by Ontario, does not affect the main results. It seems that it is those factors which determine a region's comparative advantage (geographic location, climate, endowment of natural resources, etc.) which determine the steady-state distribution of income across provinces.

Disaggregation by manufacturing industry across Canada, using monthly data on gross domestic product (GDP) per employee, reveals a bi-modal or divergent pattern. However, the manufacturing sectors in the top part of the distribution are evenly spread across provinces; and, in general, those in the bottom contribute only a small proportion of provincial GDP.

The bi-modal pattern is also evident in Ontario manufacturing industry data, but not for aggregations of western and eastern provinces. This suggests that the pattern in Ontario is driving the national picture of divergence.

The different results for provincial income per capita and manufacturing

industry GDP per employee data are striking. The income per capita data—which include transfer payments—show a unimodal long-run distribution; the GDP per employee data—which exclude transfer payments—exhibit divergence or a bi-modal distribution. This highlights the importance of transfer payments: in their absence, the long-run tendency of Canadian economic growth is towards divergence.

By ignoring the disaggregate information, existing studies that average across disaggregates present an incomplete picture of convergence. These studies fail to capture the shape and mobility dynamics within the cross-section distribution.

In the second essay, Chapter 3, I examine the impact of two types of disturbance on Canadian economic growth. Building on the finding in the previous Chapter that disaggregate dynamics contain important information for explaining the pattern of convergence, I directly identify the dynamic effects of aggregate and disaggregate disturbances. I use the long-run, just-identified structural vector autoregression (SVAR) method of Blanchard and Quah [14].

I estimate an unrestricted vector autoregression (VAR) system that includes an aggregate measure of economic growth and a measure of disaggregate interaction. To transform the unrestricted VAR to the SVAR, I assume that there exists an aggregate shock and a disaggregate or regional shock. The aggregate disturbance has permanent effects on aggregate income. This could be, for example, a technology shock. The disaggregate disturbance is identified as having no long-run impact on the level of aggregate income. For example, a provincial government fiscal stimulus might have short-run aggregate and disaggregate effects, but in the long-run leave aggregate economic activity unchanged. The two types of disturbance are assumed to be orthogonal.

Interaction, or mobility, is assumed to be unaffected by either disturbance in the long run. If disturbances had long-term effects, it would imply that the level of interaction could drift over time. However, as shown in Chapter 2, this is not the case. The short-run dynamics of the two shocks on interaction are unrestricted.

I examine annual provincial and monthly manufacturing industry data. I find that the aggregate shock has a large positive impact on provincial income per capita in both the short and the long run; and accounts for most of its variation. The disaggregate disturbance initially increases income per capita; but the

economy quickly adjusts. The effect of this disturbance becomes insignificant within 5 years. The short-run impact of the two types of disturbance on the provincial interaction measure is imprecisely estimated. The disaggregate shock provides the strongest impulse; and explains most of the variation. But its effect decays to zero within 3 years. The aggregate shock has a larger medium-term impact, which disappears within 7 years.

A similar pattern emerges with the industry data. The aggregate disturbance has a positive long-term effect on the aggregate measure, and explains most of its variation. The effect of the disaggregate disturbance is close to zero. Neither disturbance has much of an effect on the interaction measure. Within only a few months, the impacts are indistinguishable from zero.

The results in this Chapter confirm that disaggregate dynamics contain important information for explaining the pattern of economic growth, at least at business cycle horizons. By ignoring disaggregate disturbances, researchers ascribe too much importance to aggregate disturbances in explaining this pattern. The role of the aggregate disturbance in explaining most of aggregate economic activity is consistent with the neoclassical growth model, where only technological shocks are assumed to bring about long-run changes in output.

In the third essay, Chapter 4, I present the results from a variety of unit root tests on Canadian income panel data. I compare the results of the standard Dickey-Fuller method with those from three tests designed specifically for panels. Applying standard unit root tests to panel data loses information, since data must be aggregated across cross-sections. More powerful tests, which allow cross-section, as well as time-series information to be included, have recently been developed. Unlike the Dickey-Fuller test, these have limiting normal distributions.

The first panel data test is similar to the standard approach, but the initial observation for each disaggregate time-series is subtracted from each subsequent observation in that series to allow for fixed effects. The unit root hypothesis is then tested by applying a conventional t -test to the coefficient on the lagged dependent variable, using data which is averaged across the cross-section. I refer to this as the "panel Dickey-Fuller" test; see Breitung and Meyer [16].

In the second test, the data are adjusted by subtracting the cross-section mean from all observations. The test statistic is an average of individual co-

efficients on the lagged dependent variable. I refer to this as the “ δ -bar” test (Levin and Lin [51]). I extend this methodology to allow for potentially more complex common effects across individual series. Using a dynamic index model developed by Quah and Sargent [76], I identify multiple common factors and subtract the resulting indices (in place of the cross-section average) from the data.

In the third approach, the “ t -bar” test (Im, Pesaran and Shin [42]), the initial observation is subtracted—in a similar manner to the panel Dickey-Fuller case. The test statistic is a (small sample adjusted) average of the Dickey-Fuller tests for the individual series.

I examine Canadian annual data on gross provincial product and monthly data on gross domestic product per employee for 19 manufacturing industries. Using the standard Dickey-Fuller test, applied to data averaged across the cross-section, I cannot reject the null hypothesis of a unit root in either the provincial or industrial data. In contrast, none of the three panel unit root tests suggest that the data are characterised by a unit root.

This suggests that discarding cross-section information can mislead the researcher on the dynamic properties of the model. This result is in the spirit of Granger’s [36] hypothesis, in finding that aggregating a dynamic multivariate process can lead to a univariate process with fundamentally different properties.

In the final chapter, Chapter 5, I draw some conclusions.

Chapter 2

The Evolving Distribution of Canadian Economic Growth

2.1 Introduction

Canadian economic growth has long concerned both economists and politicians. Recent empirical work has focused on estimating the “rate of convergence” across provinces. Coulombe and Lee [21] and [47], Helliwell [40] and Sala-i-Martin [77] regress growth rates on initial income per capita levels. This technique was developed by Barro [3] and Barro and Sala-i-Martin [4] for looking at convergence across U.S. states. Most studies find convergence of around 1-3% per annum for various post-WWI samples. An interpretation of this result takes the form of evidence that, in the very long run, poor economies “catch up” rich economies by growing faster.

Unfortunately, this approach reveals little about the evolving pattern of economic growth within the cross-section. A convergence rate of, say, 2% per annum across Canadian provinces neither sheds light on whether rich and poor economies possess different short-run growth dynamics nor does it provide any information about the probability of a poor province improving its rank in the cross-section. An average convergence rate cannot tell us whether it is the already rich which are converging towards each other or whether the poor get (relatively) poorer. And arguably, policy makers and individuals are more likely to be concerned with the position of a province relative to others, rather than with absolute changes over time.

Another issue existing studies neglect is the relationship between provincial and industrial patterns of growth. Clearly, the two are closely related. Suppose that the pattern of economic growth for the major manufacturing industries in Newfoundland and Alberta—food and refined petroleum products respectively—is very different. Then, the finding that income per capita in the two provinces is converging should be treated with some caution. Since the

growth patterns of the leading industries differ, the path of income per capita growth is also likely to differ.

In this Chapter, I study the relationships among Canadian provincial and industrial disaggregates, explicitly modelling the dynamics. I examine both the shape of the cross-section distribution over time and the degree of mobility within it. The approach I follow builds on work by Quah [69]. I use disaggregate data on provincial income per capita and GDP per employee for manufacturing industries, to give a more complete picture of the mechanisms driving economic growth. I also examine the sensitivity of the results to different measures of economic growth (in Appendix C), to the inclusion of Ontario data, and to asymmetries in province size.

I show that although the cross-section distribution of provincial income per capita is generally unimodal, there is virtually no mobility in the cross-sectional ordering of income levels. Put differently, poor provinces remain poor and rich provinces remain rich. Newfoundland remains at the bottom of the distribution and Ontario maintains its place at the top. Adjusting the data to account for the size differences of the provinces, and for the disproportionate role played by Ontario does not affect the main results. These findings suggest that the initial distribution of resource endowments is an important determinant of the steady-state distribution of income across Canadian provinces.

In contrast to the unimodal provincial distribution, disaggregation by manufacturing industry across Canada, using monthly data on GDP per employee, reveals a bi-modal or divergent pattern. However, the manufacturing sectors in the top part of the distribution are evenly spread across provinces; and those in the bottom contribute a small proportion of provincial GDP. The bi-modal pattern is also evident in Ontario manufacturing industry data, but not for aggregations of western and eastern provinces. This suggests that the pattern in Ontario could be driving the national picture of divergence.

The different results for the provincial and manufacturing industry data are striking. The provincial income per capita data—which include transfer payments—show a unimodal long run distribution; the manufacturing industry GDP per employee data—which exclude transfer payments—exhibit divergence or a bi-modal distribution. This suggests that transfer payments play an important role: in their absence, the long-run tendency of Canadian economic growth

is likely to be towards divergence.

For policy makers concerned with provincial inequalities, this study provides a mixture of good and bad news. On the one hand, there is evidence of provincial unimodality—consistent with long-run convergence. On the other hand, there is no evidence to suggest that poorer provinces can overtake richer regions in the short or medium run.

The remainder of this Chapter is organized as follows. In Section 2.2, I review the related literature on economic convergence. In Section 2.3, I present the econometric model. I examine cross-section mobility dynamics in Section 2.4, using provincial measures of personal income per capita, and wages and salaries per employee; and manufacturing industry data on GDP per employee. In Section 2.5, I investigate the robustness of the results. I draw some conclusions in Section 2.6.

2.2 Related Literature

Following Barro [3], Barro and Sala-i-Martin [4] and [5] and Mankiw et al [55], empirical studies of economic growth and convergence have tended to concentrate on estimating rates of convergence across economies. The neoclassical growth model (Cass [19], Koopmans [44] and Solow [80]) predicts that income per capita converges to a steady-state, where per capita output, capital stock and consumption grow at a given rate of technological progress. Diminishing returns to capital imply that additions to the capital stock generate more output when the initial stock is relatively small.

Researchers begin by assuming a Cobb-Douglas production function,

$$(2.1) \quad Y(t) = K(t)^\alpha (A(t)L(t))^{1-\alpha}$$

where $0 < \alpha < 1$, Y is output, K is capital, L is labour and A is labour-augmenting technological progress. L and A are assumed to grow at rates n and g respectively, with laws of motion given by,

$$L(t) = L(0)e^{nt} \quad A(t) = A(0)e^{gt}.$$

Defining s to be the rate of saving, the dynamic equation for the stock of capital per unit of effective labour, $\hat{k}$, is given by,

$$(2.2) \quad \dot{\hat{k}}(t) = s\hat{y}(t) - (n + g + \delta)\hat{k}(t) = s\hat{k}(t)^\alpha - (n + g + \delta)\hat{k}(t),$$

where $\hat{y}$ is output per unit of effective labour and δ is the rate of depreciation. k converges to its steady state value given by,

$$\hat{k}^* = \left(\frac{s}{n + g + \delta} \right)^{\frac{1}{1-\alpha}}.$$

Substitution gives the following expression for steady state per capita income,

$$(2.3) \log(y) = \log A(0) + gt + \frac{\alpha}{1-\alpha} \log(s) - \frac{\alpha}{1-\alpha} \log(n + g + \delta)$$

where $y(t) = Y(t)/L(t)$ is income per capita. Given differing resource endowments, climate and institutions across countries, researchers assume that $\log A(0) = a + \epsilon$ where a is a constant and ϵ is a country-specific shock term.

Approximating around the steady state level of income per effective worker, $\hat{y}^*$, gives,

$$(2.4) \dot{\hat{y}}(t) = \beta[\log(\hat{y}^*) - \log(\hat{y}(t))].$$

where $\beta = (n + g + \delta)(1 - \alpha)$. Solving this equation and rearranging gives the average growth rate of per capita income y between dates 0 and T ,

$$(2.5) \frac{1}{T} \log \left[\frac{y(T)}{y(0)} \right] = g + \frac{1 - e^{-\beta T}}{T} \log \left[\frac{\hat{y}^*}{\hat{y}(0)} \right]$$

where β is the rate of convergence. If $\beta > 0$, poor economies grow faster than rich economies—as predicted by neoclassical growth models. This is known as *β -convergence*. Defining σ^2 to be the variance of y , the data show *σ -convergence* if $\sigma_t \leq \sigma_{t-1}$ for all t . That is, there is a decline in dispersion, over time, in the (entire) cross-section.

Researchers estimate equations similar to equation (2.5). The coefficient on $\log \left[\frac{\hat{y}^*}{\hat{y}(0)} \right]$ declines in magnitude with the length of the sample, for a given β . As T increases, the effect of the initial position on the average growth rate gets smaller; as $T \rightarrow \infty$ the coefficient tends to zero. β is estimated non-linearly to take account of the associated value of T in equation (2.5). In this way, similar estimates of β should be obtained, regardless of the length of the sample.

Estimated on a wide range of data—across and within countries—values of β derived from this model suggest a uniform convergence rate of 2% per annum. For example, Barro and Sala-i-Martin [5] estimate β at 0.0175 (standard error 0.0046) for per capita personal income in the US states over the period 1880–1988. In an earlier paper ([4]), the authors show that the dispersion of US state

personal income has declined from about 0.50 in 1880 to around 0.20 in the 1980s.

In [4], Barro and Sala-i-Martin also examine convergence across U.S. non-agricultural sectors over the period 1963-1986.¹ Estimates of β range from 0.0093 in Wholesale and Retail Trade, to 0.0460 in Manufacturing (standard errors are 0.0064 and 0.0082, respectively).

Barro and Sala-i-Martin [5] obtain similar results from cross-country studies. For a sample of 98 countries, they find evidence of “conditional” convergence for the period 1960 to 1985. The neoclassical growth model implies only that economies with the same preferences and technology will reach a common steady state. When countries are diverse, it is necessary to allow for heterogeneity in certain institutional variables—for example, initial school enrolment rates, the ratio of government consumption expenditure to GDP, and proxies for political stability—in order to apply the framework in cross-country studies. This is the sense in which the convergence is “conditional”.

Several studies recognise that the transfer of technology is an important part of the story. Using annual data (1963-1989) on technical progress in OECD countries, Helliwell [39] shows that there has been significant international convergence in the rate of technical progress: initially poorer countries have faster technical progress. In Helliwell and Chung [41], analysis of the Solow residuals for OECD countries for 1960-1985 shows convergence in rates of technical progress. The authors conclude that convergence in per capita GDP is not just a function of differences in investment rates.

Recent Canadian studies include Coulombe and Lee [21] and [47], Helliwell [40], Lee [46] and Sala-i-Martin [77]. Coulombe and Lee ([21] and [47]) examine Canadian provincial income and output per capita data during 1961-1991. They split the sample into three sub-periods, 1961-1971, 1971-1981 and 1981-1991, and regress each province’s growth rate, relative to the Canadian average, on the initial level of income, relative to the Canadian average, for each time period. This approach helps to mitigate the effect of sample-specific time trends. Coulombe and Lee report estimates of β that range from 0.0105 for gross provincial product per capita to 0.0289 for personal disposable income

¹The eight sectors are: mining; construction; manufacturing; transportation; wholesale and retail trade; finance, insurance and real estate; services; and government.

per capita.

Lee [46] uses an extended version of the neoclassical model to account for differences in various investments and industrial structures across Canadian provinces. Annual growth in output per worker is regressed on the initial level of output per worker, the sum of technical progress, labour growth and the capital depreciation rate, and the different investment rates (human, private physical and public physical). Over the period 1966-1992, Lee [46] finds that the dispersion of productivity declined steadily.

Helliwell [40] shows that provinces which were poorer in 1961 have faster average growth rates over the period 1961-1989 and that there is a significant downward trend in the interprovincial variation of income levels during 1926-1990. Sala-i-Martin [77] estimates β for Canadian income per capita between 1961 and 1991 at 0.024 (standard error of 0.008).

Traditional methods of measuring convergence are also related to growth patterns among trading partners (Ben-David [8]) and in explaining educational and financial development (Berthelemy and Varoudakis [11]). Ben-David [8] finds that grouping countries according to their primary trade affiliations tends to produce significant income convergence within the groups. Such significant convergence is uncommon among these countries when they are grouped randomly. Berthelemy and Varoudakis [11] find that two groups of countries, separated by a financial development threshold, form distinct convergence clubs to different steady-state growth paths.

Several recent papers question the methodology used in the traditional approach. Evans and Karras [32], for example, show that the traditional approach to estimating convergence is only valid if economies have identical first-order autoregressive dynamic structures and all permanent cross-country differences are completely controlled for. They develop an alternative approach, based on Levin and Lin's [51] unit root test.² However, the traditional approach and Evans and Karras' [32] methodology lead to similar conclusions: the 48 contiguous US states and a group of 54 countries show conditional convergence.

Lee, Pesaran and Smith [48] consider the predictions of the Solow theory using panel data on output per capita for 102 countries for the period 1960-1989. They examine three notions of convergence: β -convergence; σ -convergence; and

²See Chapter 4 for more details of this test.

whether, in the time-series dimension, each country is converging to its own steady-state equilibrium, i.e. whether the data contain a unit root. They strongly reject a common steady-state growth rate: across the sample, the variance of the log of per capita output increases with time. Time-series estimates of the speed of convergence are about 20%, but it is argued that these are biased upwards. This is confirmed by a range of unit root tests which suggest that the series are non-stationary. The authors cannot reject the hypothesis of no convergence, even to country-specific steady states.

Unfortunately, the estimation methods used by all these studies ignore the evolving pattern of the cross-section distribution. No account is taken of the nature of interaction across economies. As Quah ([67] and [75]) argues, β -convergence, or looking at the coefficients of a cross-section regression, gives no information about the changing nature of that distribution over time. These coefficients represent only average behaviour. A convergence rate neither provides information about transitory growth dynamics across economies nor does it indicate anything about the relative rankings of different provinces. Conditional convergence only shows whether each economy converges to its own steady state, which is different from that of other countries.³ Similarly, a measure of cross-section variance (σ -convergence) cannot distinguish between a situation in which economies are criss-crossing within the cross-section and a situation in which the rich remain rich and the poor remain poor.

Suppose the income distribution across economies can be represented in time t as shown by the density distribution in Figure 1. This shows that the majority of economies are in the middle of the distribution. For some time $t + s$, $s > 0$, another density distribution is shown. This is bi-modal: rich economies group together; poor economies group together; and the middle ground is disappearing. The figure illustrates both the external “shape” of the cross-section distribution, and the internal “mobility” or churning within the distribution. Traditional convergence studies do not address either of these data characteristics. But these dynamics provide important information on the dynamics of the poor catching up the rich and on convergence clubs.

³Conditioning on explanatory variables leads researchers to conclude that it is these variables that determine an economy’s position. But, as pointed out by Quah [70], it is the factors determining “club” membership that are important. When different convergence clubs form, factor inputs and social characteristics endogenously align around values determined by each economy’s convergence club.

Quah [69] develops an empirical model that incorporates mobility and shape dynamics. The econometric method uses information from the entire distribution, and imposes no structure, on either growth trends or the expected nature of convergence. Using this model enables investigation of short-run growth dynamics. It can, for example, yield information on whether there are distinct, identifiable groups (clubs) of rich and poor. Within the cross-section distribution, the nature of rankings of observations can also be examined, such as whether provinces which were relatively poor yesterday remain so today.

Quah [64] uses this model to investigate movements in GDP per capita income in 118 countries, relative to the world average, over the sample period 1962-63 to 1984-85. Steady-state distributions suggest cross-country incomes tend towards extremes of the cross-section. That is, middle-income countries are disappearing in favour of the very rich and very poor. In a later paper, Quah [75] studies mobility across US states—the results suggest convergence, or a unimodal steady-state. However, persistence remains high at the ends of the cross-section distribution.

Using similar methods to Quah, Andres and Lamo [1] analyse GDP per capita data for 24 OECD economies from 1960 to 1990. They find evidence of inertia in income rankings and one group of economies remains persistently at the bottom of the cross-section distribution.⁴

Bianchi [12]) studies data on GDP per capita for 119 countries in 1970, 1980 and 1989. He takes each cross-economy income distribution at time t in isolation and estimates it non-parametrically. He then applies a bootstrap test of multimodality to each. The data show unimodality in the early part of the sample (early 1960s); but by the end of the sample (late 1980s), the data reject unimodality in favour of bimodality.

In this paper, I use the techniques developed by Quah [69] to examine Canadian provincial and manufacturing industry data. By considering both provincial and industrial data, I gain a more complete picture of the factors driving the pattern of economic growth and convergence. In an early study of regional aspects of Canada's economic growth, Green [37] finds that, in 1956, high income

⁴The bottom 25% in 1960 fell in a range of 0.26 to 0.54 of the OECD average. By 1990, 25% of OECD countries still had incomes between 0.3 and 0.8 of the OECD average. By contrast, countries in the second and third quantiles fluctuated about the mean, while the richest 25% started with an upper limit of 1.56 and ended at 1.25 times the average.

provinces tended to have either a large manufacturing sector (where output per worker is high) or a large share of output concentrated in a relatively high productivity resource output. He concludes that not only was the distribution of resources between agricultural and non-agricultural output important, but the composition and productivity within these sectors appeared to be important determinants of income differences.

2.3 Econometric Model

In this Section, I outline the econometric framework developed by Quah (see, for example, [69]), which I subsequently use to examine the evolving nature of the distribution of economic growth across Canada.

First I assume that the cross-sectional distribution of the n -dimensional vector of disaggregate variables, X_t , is generated by disaggregate-specific shocks, conditional on aggregate effects.⁵ Denote the cross-sectional distribution of disaggregates at time t by $\mathcal{F}_t$. Then, following Quah [69], I employ a stochastic kernel equation to represent the dynamics of the cross-section distribution.⁶ The dynamics of the distribution in $\mathcal{F}_t$ are parameterized by (M, Q) , where M is a transition probability matrix and Q is a sequence of quantiles. Each quantile-set pair $(Q(t), Q(t+1))$ defines a transition probability matrix M of transitions from $\mathcal{F}_t$ to $\mathcal{F}_{t+1}$. The matrix M summarizes information on cross-section mobility; the sequence Q describes the shape of the cross-section distribution.

Each element (j, k) of the transition probability matrix, M , indicates the probability of an observation starting in state j in period t , ending up in state k in period $t+1$. The probability of arriving at state j in period t depends only on the state the variable was in one period earlier—the history of arriving at a

⁵These disaggregate shocks have only transitory effects on the aggregate. They could include, for example, a regional demand shock, or a drop in fish stocks or a bumper harvest in the prairie provinces.

⁶A kernel function K is usually a symmetric probability density function, such as the normal density. The kernel estimator, with kernel K is defined by,

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

where X_i is a random variable, and h is the window width (or smoothing parameter). The kernel estimator is a series of “bumps” placed at the observations. The kernel function K determines the shape of the bumps while the window width h determines their width. In this estimation, a squared Epanechnikov kernel is used (see Silverman [79]).

certain point is ignored.⁷

S -period ahead transition probabilities are obtained by multiplying the transition probability matrix M by itself s -times (i.e., M^s). Varying the transition period indicates how mobility in the cross-section distribution behaves over time. These s -period transition probabilities contain information about both short and long-run disturbances. As s changes, the model imposes identifying restrictions in different ways on the transition probability matrix, M . As a consequence, the information contained in the transition probability matrix is a function of the transition period. Very short periods provide details of short-run fluctuations, but are relatively uninformative for long-run shocks. As s increases, a more accurate picture of the impact of long-run disturbances is provided.⁸ I present results for different transition periods, enabling a comparison of the relative importance of short and long-run shocks to the data.

I also compute the ergodic or steady-state distribution implied by the transition probability matrix M . This is given by the eigenvector associated with the (single) unit eigenvalue.⁹ If it exists, the ergodic distribution provides a summary of intradistribution mobility—with aggregate effects removed, it represents movements of disaggregates relative to each other.

2.4 Cross-Section Dynamics

In this Section, I investigate the shape of and mobility in the distribution of disaggregate measures of Canadian economic growth.¹⁰ I examine disaggregation by province and by manufacturing sector. Appendix A lists the sources of data.

⁷More precisely, suppose X_t is a random variable that can assume only an integer value $\{1, 2, \dots, N\}$, then,

$$P(X_t = k | X_{t-1} = j, X_{t-2} = i, \dots) = P(X_t = k | X_{t-1} = j) = p_{jk}.$$

This is an N -state Markov chain with transition probability p_{jk} . Markov chains form the simplest time-series model for a discrete-valued random variable.

⁸This must be weighed against the accompanying loss in “power”. In estimating s -year transition probabilities, the true sample size is approximately T divided by s . That is, the number of non-overlapping samples helps to determine the power of the test. As s increases, the number of non-overlapping samples decreases.

⁹A distribution is *ergodic* if one of the eigenvalues of the transition probability matrix M is unity, and all other eigenvalues are inside the unit circle. The long-run forecast for an ergodic Markov chain is independent of the current state. That is, the vector of ergodic probabilities indicates the unconditional probability of each of the different states.

¹⁰Estimation in this thesis is carried out using SHAZAM, RATS and Danny Quah’s time series random fields (tsrf) package.

1-year transitions					
(Number)	Upper endpoint				
	-0.365	-0.207	-0.071	0.074	0.290
(130)	0.90	0.09		0.01	
(129)	0.05	0.83	0.10	0.02	
(125)	0.01	0.06	0.77	0.14	0.02
(126)	0.01	0.02	0.16	0.76	0.05
(128)				0.09	0.91
Steady-state distribution	0.149	0.205	0.245	0.229	0.172
5-year transitions					
(Number)	Upper endpoint				
	-0.367	-0.215	-0.071	0.074	0.290
(128)	0.78	0.20	0.01	0.02	
(115)	0.03	0.76	0.15	0.05	0.01
(111)	0.02	0.06	0.70	0.20	0.02
(113)	0.02	0.03	0.22	0.62	0.12
(121)		0.02		0.17	0.82
Steady-state distribution	0.073	0.175	0.285	0.264	0.204

Table 2.1: Transition Probability Matrix—Personal Income per Capita by Province

2.4.1 Provincial

I look at two different measures of economic growth at the provincial level of disaggregation: personal income per capita, and wages and salaries per employee.

Personal income

Figure 2 plots annual real personal income per capita, relative to the national average, for the Canadian provinces over the period 1926-1993. In 1926 PEI's real income per capita was 33% below the national average; by 1993, the difference was only 10%. In contrast, BC's real income per capita exceeded the national average by more than 46% in 1926; by 1993, this figure had fallen to 15%. The shape of the cross-section distribution has also altered, with the dispersion of incomes declining markedly over the sample.

Using the (log) ratio of provincial real per capita income to the national average, I estimate a transition probability matrix, M , following the methods outlined in Section 2.3. Table 2.1 presents the results for two different assumptions about the period of transition. The endpoints of the intervals (quantiles) across

the cross-section are designed to ensure that approximately equal numbers of observations begin in each interval.¹¹ The upper endpoints of each interval are shown on the top row of each panel. The first column gives the total number of observations with starting points in that income state.¹² I present the results for $n = 5$, where n is the number of quantiles.¹³ Recall that each (j, k) entry of the transition matrix is the probability of a province in state (or quantile) j in period t transiting to state k in period $t + s$.

The first panel of Table 2.1 contains the annual transition matrix, estimated by averaging the observed one-year transitions over every year, from 1927 to 1993.¹⁴ This shows high persistence for the lower and upper quantiles: the corner entries are at least 90%. The ergodic distribution implied by the one-year transition function shows an accumulation in the middle states, and a thinning at the upper and lower tails.¹⁵

The second panel of Table 2.1 shows the transition probability matrix produced from estimation with a five-year transition period.¹⁶ As expected, mobility increases relative to the one-year transition period: diagonal elements are smaller, and off-diagonals larger. The ergodic distribution becomes more skewed—away from the lower states and towards the upper states (3 and 4)—but remains unimodal.¹⁷ The changing shape of the steady-state distribution provides an indication of the errors in the short transition period estimates. A

¹¹The intervals are constructed from uniformly distributing the data over its observed range. An alternative approach would be to specify the end-points of each interval, but this would impose (possibly incorrect) structure on the data.

¹²Where numbers in the first column differ from each other, it is by at most the number of years in the sample. This is due to rounding errors.

¹³I experimented with n ranging from 2 to 6, but found the results varied little.

¹⁴Entries in the lower right-hand section of each panel in Table 2.1 show the probability of an economy in a rich state staying in that state in the next period. Entries showing 0 to 2 decimal places are left blank.

¹⁵Estimation of a second-order transition matrix ($X(t)$ depends on $X(t - 1)$ and $X(t - 2)$) produces a similar distribution in the steady state. In this case, the corner diagonal entries of M are both 0.95.

¹⁶The entries in each column of each panel should sum to one. Although this holds for the one-year transition period (given minor rounding errors), it does not always apply in the longer transition periods. One possible explanation for this discrepancy lies with the sparse nature of the transition probability matrix—there are a non-trivial number of zero entries. I investigate this by experimenting with several different matrices, and find that several zero entries lead to summation errors. With no zero entries, the sum of column entries is equal to one, even for 20-year transition periods. With zero entries, errors begin to appear after only a few years.

¹⁷Estimation with a ten-year transition period produces similar results, with slightly increased mobility over the five-year transition period, but a similar ergodic distribution. I also restrict the sample to the period 1961-1992. Mobility and ergodic distributions vary little from the results based on the longer sample period.

1-year transitions					
(Number)	Quantile				
	0.20	0.40	0.60	0.80	1.00
(109)	0.92	0.08			
(131)	0.07	0.77	0.15	0.02	
(130)		0.15	0.73	0.12	
(129)		0.02	0.12	0.84	0.02
(130)				0.02	0.98
1-year transitions					
(Number)	Quantile				
	0.25	0.50	0.75	1.00	
(130)	0.93	0.06	0.01		
(174)	0.05	0.81	0.14	0.01	
(130)		0.20	0.68	0.12	
(195)			0.08	0.92	

Table 2.2: Fractile Transition Probability Matrix—Personal Income per Capita by Province

one-year transition period is too short to provide an accurate indication of the steady-state outcome. However, it yields similar qualitative results to the longer transition periods.

An alternative method of estimating transition probabilities is to examine the time-invariant fractile matrix and the sequence of quantiles.¹⁸ A fractile matrix possesses elements in which every quantile set has equal measure. When M is fractile, the uniform distribution is always an ergodic limit. Table 2.2 presents fractile transition probability matrices for provincial income per capita data for two different quantile specifications (which are shown in the top row of each panel). These matrices are approximately the average of a sequence of time-varying fractile transition probabilities. In comparison with the top panel of Table 2.1, the corner entries exhibit more persistence, particularly at the top of the distribution. However, mobility between states appears more pronounced at the lower end of the distribution.

¹⁸ A matrix $M(t)$ is fractile, if,

$$\left(\sum_{m=1}^n M_{lm}(t) \right) \phi_{y,t}(q_l(t)) = \phi_{y,t}(q_l(t)) = \phi_{y,t}(q_1(t)) \quad \forall l.$$

Differences between the matrices in Table 2.1 and Table 2.2 can be attributed to the different methods of estimation. In Table 2.1, the transition probability matrix imposes a uniform distribution at the outset, but loses precision at the ends of the distribution. By contrast, the fractile matrix forces the data to be uniform in the limit, pushing the mid-lower and mid-higher income groups into the tails of the distribution. This characteristic also helps to explain the relatively large entries in quantiles 2, 3 and 4 in the off-diagonal (Table 2.2): they reflect a tendency for the distribution to become more uniform over time.

Figure 3, a chart of the quantiles (associated with Table 2.1), reveals that dispersion and variance across the cross-section is declining. The chart plots the log of provincial income per capita in quantile i over average income per capita against year. The bottom two quantiles display relatively steep gradients, whereas the top quantile appears much flatter: economies in the bottom of the cross-section distribution have moved up the distribution faster than those at the top have moved down—the gap between richest and poorest is narrowing. However, the ranking of provinces within the cross-section distribution remains largely unchanged over the sample—poor provinces (Newfoundland and PEI) remain at the bottom of the distribution, rich provinces (Ontario and Alberta) remain at the top. It is this persistence at the ends of the cross-section distribution which is ignored in the traditional studies of convergence.¹⁹

Wages, salaries and supplementary labour income

I now study the distribution of wages and salaries per employee across provinces.²⁰ Annual data on wages, salaries and supplementary labour income (WSSLI) are available for the period 1966-1994.²¹ I examine WSSLI per employee, relative to the national average. Fluctuations in the data are concentrated around the early 1980s (Figure 4). Unlike the income per capita data,

¹⁹ Appendix C contains results for personal income per capita growth data. The transition probability matrices show mobility is higher than in the levels data, and there is no pattern of persistence. The ergodic distributions are unimodal and symmetric. I also study data on provincial annual employment per capita for 1966-1992: cross-section dispersion changes little over the sample period, and mobility within the distribution is limited. The ergodic distributions suggest convergence towards the bottom half of the cross-section distribution.

²⁰ Appendix C contains results for estimation of transition probability matrices using data on provincial wages and salaries per capita. The pattern of mobility is very similar to that found in the personal income per capita data: persistence at the ends of the cross-section distribution, and a unimodal ergodic pattern.

²¹ WSSLI, unlike personal income, excludes interest and dividend earnings.

1-year transitions					
(Number)	Upper endpoint				
	-0.112	-0.023	0.019	0.118	0.279
(54)	1.00				
(54)		0.78	0.22		
(54)		0.20	0.74	0.06	
(54)			0.06	0.80	0.15
(54)				0.15	0.85
Steady-state distribution	0.264	0.052	0.057	0.264	0.363
3-year transitions					
(Number)	Upper endpoint				
	-0.112	-0.023	0.019	0.118	0.279
(48)	1.00				
(49)		0.65	0.33	0.02	
(47)		0.32	0.55	0.13	
(47)			0.15	0.64	0.21
(49)				0.22	0.78
Steady-state distribution	0.226	0.073	0.079	0.238	0.384

Table 2.3: Transition Probability Matrix—Wages, Salaries and Supplementary Labour Income per Employee, by Province

Newfoundland now fluctuates about the average rather than remaining at the bottom of the distribution, reflecting the relatively high unemployment rate in Newfoundland.²²

The transition probability matrices (Table 2.3) show that economies starting in the bottom quantile are unlikely to move up the distribution.²³ However, there is a very high probability of leaving the middle states, particularly with a 3-year transition period. The ergodic distributions display a very different pattern from the per capita data. In this case, economies tend to diverge to the ends of the cross-section distribution.²⁴

Three main groupings are visible in the quantiles generated from the WSSLI per employee data (Figure 5). The bottom group displays an upward drift; the

²²In 1976, Newfoundland's unemployment rate averaged 13.1%; the average across Canada was 7.2%. By 1994, Newfoundland's unemployment rate had risen over 50% to 20.4%, while the national average had risen only 45% to 10.4%. Newfoundland's unemployment rate remained the highest of the ten provinces over the period 1976-1994.

²³Fractile transition probability matrices for this and all subsequent variables studied in this Section are in Appendix B.

²⁴The differences are not attributable to the different sample periods: re-estimation using data for 1966-1992 for WSSLI per capita and WSSLI per employee did not produce significantly different results.

middle group is fairly flat (centred about zero); and the top group exhibits a slow decline. Dispersion has fallen over time. This suggests convergence in the traditional sense: a narrowing of the cross-section range over time. However, the ergodic distributions suggest divergence *within* the cross-section. Although the range across provinces of WSSLI per employee is falling, provinces with relatively low WSSLI tend to remain at the bottom of the distribution; those with a relatively high value stay at the top. This points to a bimodal pattern. In the middle of the cross-section distribution, the mass is approximately zero.

Differences in the ergodic distributions between per capita and per employee data can be attributed to the gap between the number employed and the total population in the different provinces. In general, those provinces with low population have a high unemployment rate; those with high population have a lower unemployment rate. Migration to the richer provinces may have increased the gap between rich and poor (in terms of income and employment opportunities). This is consistent with Helliwell [40], who finds that income per capita and unemployment rate gaps are strong determinants of migration²⁵ and Lee [46], who finds that inter-provincial migration appears to slow down the convergence process.

Lee and Coulombe [47] find that output per worker converges faster than output per capita, and attribute this to regional disparities in unemployment rates.²⁶ Although I find that although dispersion has fallen, WSSLI per employee diverges within the cross-section distribution. The discrepancy between relatively rich and poor provinces falls. However, at the same time, the probability that the relatively poor can become rich shrinks.

2.4.2 Industrial

I now consider disaggregation by manufacturing group.²⁷ Combining the results from this Section with those for the provincial disaggregation provides a clearer

²⁵ Migration has primarily been directed towards the richer provinces.

²⁶ Appendix E contains results from estimations using data employed by Coulombe and Lee in their studies of convergence across Canadian provinces (see [21], [47] and [22]). In most cases, my results are consistent with the reduction in dispersion of different measures of output and income found by Coulombe and Lee. However, cross-section dynamics differ across variables, and for one measure—output per hour worked—there appears to be divergence within the distribution.

²⁷ Appendix C contains results from estimation for disaggregation by seven industry sectors, examining GDP per person at work and GDP per person hour for the period 1961-1994. Both data sets produce unimodal ergodic distributions.

picture of the evolving pattern of economic growth. It enables an analysis of whether particular industries can explain the pattern of provincial income.

I study gross domestic product (GDP) per employee (relative to the national average) for 21 manufacturing industries, using monthly data for the period 1961:1 to 1994:11.²⁸ GDP per employee varies both across industries and over time, as illustrated in Figure 6. The tobacco and tobacco products industry has the highest GDP per employee, followed by the refined petroleum and coal products, and beverages industries. At the bottom of the distribution are the leather and allied products, clothing, and furniture industries. The refined petroleum and coal products industry possesses the highest value of capital stock per employee; the lowest value is held by the clothing industry. GDP per employee may, to a large extent, be reflecting the capital intensity of the different industries. Alternatively, it may be an indication of the low value-added in or productivity of those industries at the bottom of the distribution.

Table 2.4 presents the transition probability matrices. One-month transition period probabilities show surprising amounts of movement among states, particularly in the middle ranges. However, industries do not move further than one state away from their initial position, and in the middle ranges, are equally likely to move up or down. The ends of the cross-section distribution show more persistence. As the transition period is increased to three years, mobility increases across the board.

The steady-state distributions reveal an interesting pattern. In the one-month transition period, the steady-state is unimodal. As the transition period increases, a bi-modal pattern becomes evident—industries tend to group at the bottom or top of the cross-section distribution. At first glance, this result seems to contradict the provincial results which reveal that the ergodic distribution is unimodal. However, the manufacturing industries which are at the bottom of the cross-section distribution (leather, clothing and furniture and fixtures) do not, in general, contribute a great deal to provincial GDP. Those industries at the top of the distribution (refined petroleum and coal products, tobacco, beverages and food) are evenly spread across provinces. This suggests that, at least in the long run, there is some support for the traditional theory of convergence.

²⁸ Appendix A lists the manufacturing industries used in the estimation.

1-month transitions					
(Number)	Upper endpoint				
	-0.594	-0.202	0.068	0.325	1.140
(1648)	0.94	0.06			
(1648)	0.06	0.85	0.09		
(1651)		0.09	0.80	0.10	
(1649)			0.10	0.81	0.09
(1646)				0.09	0.90
Steady-state distribution	0.202	0.197	0.199	0.201	0.201
6-month transitions					
(Number)	Upper endpoint				
	-0.594	-0.202	0.068	0.325	1.140
(1624)	0.92	0.08			
(1623)	0.09	0.79	0.12		
(1634)		0.11	0.71	0.17	0.01
(1629)			0.17	0.67	0.16
(1627)			0.01	0.16	0.83
Steady-state distribution	0.205	0.194	0.202	0.201	0.198
12-month transition					
(Number)	Upper endpoint				
	-0.594	-0.202	0.068	0.325	1.140
(1593)	0.92	0.08			
(1594)	0.09	0.81	0.10		
(1619)		0.09	0.79	0.11	
(1606)			0.11	0.79	0.09
(1599)				0.10	0.90
Steady-state distribution	0.214	0.188	0.202	0.201	0.195
36-month transition					
(Number)	Upper endpoint				
	-0.594	-0.202	0.068	0.325	1.140
(1371)	0.86	0.12			
(1382)	0.14	0.69	0.17		
(1435)		0.15	0.66	0.19	
(1415)			0.18	0.70	0.12
(1400)				0.13	0.87
Steady-state distribution	0.214	0.178	0.199	0.210	0.199

Table 2.4: Transition Probability Matrix—GDP per Employee, 21 Manufacturing Industries

In summary, for all disaggregates studied, there is considerable persistence at the extremes of the cross-section distribution.

I find that there has been a reduction in cross-section dispersion for measures of provincial income per capita over the period 1926-1992. Moreover, the relative increase in income per capita experienced by the provinces at the bottom of the rankings is larger in absolute value than the reduction at the top. However, there is also widespread persistence—little cross-section mobility—at the extremes of the cross-section distribution. In particular, Newfoundland remains at the bottom of the distribution; Ontario and Alberta remain at the top. By contrast, there is considerable mobility among those provinces in the middle of the distribution.

Data on WSSLI per employee reveal a different pattern. The ergodic distribution of WSSLI exhibits divergence within the cross-section. Those provinces with higher than average unemployment rates and lower than average wage rates diverge from those at the top of the cross-section distribution.

Disaggregation by manufacturing industry, using monthly data on GDP per employee from 1961 through 1994, exhibits relatively high mobility even over a one-month transition period. In this case, the ergodic distribution is bi-modal. This could reflect different technologies and varying capital intensities across manufacturing industries.

2.5 Robustness

In this Section, I study the sensitivity of the results to province size (measured in terms of contribution to national GDP). I adjust the data in different ways before estimating the transition probability matrices. First, I adjust the data for the asymmetric size of the provinces; in particular, I test for any possible bias in the results of Section 2.4 contributed by the size of Ontario relative to the other provinces. Ontario accounts for approximately 40% of national GDP and personal income, and one third of the total population.²⁹ Second, I study the impact on the results of aggregating data for some of the smaller provinces. Finally, I investigate the role of Ontario in influencing the manufacturing indus-

²⁹The shares of GDP and income attributable to Ontario dip slightly in the early 1980s. This coincides with an upturn in Alberta's share of Canadian GDP and income, as the refined petroleum and chemicals industries increase in importance in that province.

try data by aggregating the provincial data by region and comparing the results for different regions.

Appendix D contains the charts and transition probability matrices from the estimations performed in this Section.

2.5.1 Provincial data adjustment

In the first adjustment to deal with any potential bias introduced by the relatively large size of Ontario, I divide all observations by the corresponding observation for Ontario. In the second adjustment, I subtract Ontario data from the national total. These two adjustments allow a comparison of absolute and relative effects. The absolute adjustment emphasises the size differential; the relative adjustment focuses on any differences in mobility. Thirdly, given the disproportionately small size of some provinces, I aggregate data from some of the provinces to form six different geographic regions: the Maritimes (Newfoundland, PEI, Nova Scotia and New Brunswick), Quebec, Ontario, the Prairies (Manitoba and Saskatchewan), Alberta and British Columbia.

Relative to Ontario (1)

The first adjustment—dividing the data by the corresponding observation for Ontario—displays a similar pattern of persistence in the personal income levels data, but produces a slightly different steady-state pattern.

The adjustment produces smaller actual numbers—but the fluctuations follow a similar pattern.³⁰ The adjusted data exhibit more mobility, particularly in the bottom half of the cross-section distribution.

The adjusted data also suggest convergence (a reduction in dispersion) in income per capita over time, insofar as the one-year ergodic distribution shows a similar pattern to the unadjusted data—it is broadly unimodal. As the transition period increases, the unadjusted data produces a graph which peaks in state 4; the adjusted data peak in the final state. This is reflected in the transition probability matrices—there is a higher probability of moving down a state than of staying put with the adjusted data compared with the unadjusted data.

³⁰Graphing GDP growth for Ontario and Canada between 1962 and 1994 shows that peaks and troughs are largely correlated and tend to occur in the same years. The recession in Ontario in 1981 shows clearly in the adjusted data set, but is barely visible in the unadjusted data. The variance for the Canadian data is slightly higher than that for Ontario.

This is because the adjusted data is divided by data from the highest value province, rather than by the average.

Relative to Ontario (2)

In the second adjustment, I subtract Ontario data from the observations for other provinces.³¹ I find (like the first adjustment) that the steady-state distributions suggest a shifting up of the cross-section distribution as the transition period increases.

The quantiles for the unadjusted and adjusted data (see Figures 3 and D2) display very different patterns. In the unadjusted data, there is a clear tendency for the bottom quantiles to display faster growth than the others. In the adjusted data, no such pattern is discernible. Instead, fluctuations occurring around 1981 (peaks in the top two quantiles, and troughs in the bottom three) are greatly magnified. This reflects the marked dip in personal income per capita data in Ontario during the 1981 recession.

These results confirm the importance of Ontario in driving the national results. Absolute differences produce different quantile patterns, but ergodic distributions do not differ significantly from the unadjusted data.

Aggregating some provinces

In the third adjustment to the provincial income data, I aggregate observations for some of the smaller provinces to form larger regions. In this case, the results closely resemble those found in the unadjusted case. However, there is greater persistence at the bottom end of the cross-section distribution, which is reflected in the ergodic distributions: observations converge towards the middle of the cross-section distribution, rather than the top end. This is because, when aggregated, the poorer provinces have a more significant effect on the results, contributing to a downward shift in the unimodal peak.

Figure D3, which charts the different quantiles, shows that they divide into three different groups. The ranking of provinces within these quantiles is as follows: the Maritimes and the Prairies at the bottom; Quebec and Alberta in

³¹I do not convert this data into logs.

the middle; and British Columbia and Ontario at the top. Both dispersion and variability diminish over time.

2.5.2 Industry data

I present estimates of transition probability matrices for GDP per employee in 21 manufacturing industries, using three different provincial industry aggregates. The data set covers Ontario, the western provinces (B.C., Alberta, Saskatchewan and Manitoba), and the eastern provinces (Quebec, New Brunswick, Nova Scotia, PEI and Newfoundland).³² These estimations enable me to examine more closely the role played by Ontario in the industry data.

Ontario

I use annual data on GDP per employee, relative to the manufacturing industry average in Ontario, over the period 1971 to 1991. The transportation equipment industry contributes the highest value to Ontario GDP (more than 15%), followed by industries manufacturing electrical and electronic products (10%), primary and fabricated metal (9%), and food products (9%). Employee numbers follow a similar pattern. A wide range of industries is represented in the province.

The transition probability matrices (Table D4) show considerable mobility even within a one-year transition period. Once again, most persistence is seen at the ends of the cross-section distribution: observations tend to accumulate in one of two tails. Over different transition periods, the ergodic distribution is generally U-shaped.³³ This suggests that, as for manufacturing industries across Canada, GDP per employee for manufacturing industries in Ontario tends to diverge to the ends of the distribution. Some sectors, the winners, remain at the top of the cross-section distribution; others, the losers, stay at the bottom.

Figure D4, showing the quantiles, looks similar to the quantile chart for manufacturing industries across Canada. Most of the volatility appears in the top and bottom quantiles. These quantiles are also further apart from the remaining quantiles.

³²See Appendix A for a list of the industries used in the estimation.

³³A six-year transition period produced a similar pattern, with increased mobility at the bottom of the cross-section distribution.

Western provinces

Next I aggregate annual data (1971-1991) for GDP per employee in the same 21 manufacturing industries for the four western provinces of B.C., Alberta, Saskatchewan and Manitoba.³⁴ One- and three-year transition probability matrices (Table D5) show a high degree of mobility across quantiles: even over one year, most states can be reached from others. There is more persistence at the bottom of the distribution than at the top. The ergodic distributions display a largely uniform pattern—there is no tendency for GDP per employee to either converge or diverge over time.³⁵ The chart of the quantiles (Figure D5) shows that again, most volatility is contained in the end quantiles.

Eastern provinces

Finally, I consider GDP per employee for the eastern provinces—Quebec, New Brunswick, Nova Scotia, PEI and Newfoundland—relative to the eastern average for 1971-1991. The data exhibit considerable volatility. This is particularly true in the chemical and refined petroleum industries.

Estimation of one and three-year transition probability matrices (Table D6) shows that mobility at the bottom of the distribution is lower than for the western provinces, with a zero probability of moving up more than one quantile. Ergodic distributions show increasing concentration towards the top end of the cross-section distribution.³⁶ The chart of the quantiles (Figure D6) shows, once again, that it is the upper and lower quantiles which display most volatility. The degree of volatility has decreased since 1980, and this appears particularly true in the bottom quantile.

Given a few minor differences, adjusting the data to account for the size differences of the provinces, and for the disproportionate role played by Ontario, does not appear to affect the main results. Aggregate fluctuations in Canada follow a similar pattern to those seen in Ontario. In Ontario, ergodic patterns produced by estimation of transition probability matrices suggest divergence in

³⁴I weight the contribution of each province by the ratio of its population to total 'western' population.

³⁵Estimation with a six-year transition period produces an ergodic distribution which follows a similar pattern, showing a shallow peak in the middle of the distribution.

³⁶A six-year transition period produces similar results.

GDP per employee (to extremes of the cross-section distribution). By contrast, the aggregation of western provinces produces roughly uniform ergodic distributions and the eastern provinces exhibit unimodality or convergence towards the top end of the cross-section distribution. These differing ergodic distributions suggest that the pattern in Ontario may be driving the national picture of divergence.³⁷ This should not be surprising. Ontario produces a large share of national output in a wide range of manufacturing industries.³⁸

2.6 Conclusions

In this Chapter, I examine the dynamics of economic growth in Canada, investigating the evolving nature of the underlying cross-section distribution. Most previous studies of economic growth and convergence in Canada focus on aggregate measures, and conclude that convergence occurs across countries (or provinces/states) and across industries. However, the methods used reveal little about fluctuations in the pattern of economic growth within the cross-section. They tell us nothing about the mobility within the cross-section or the shape of the distribution.

Using an econometric framework developed by Quah [69] that incorporates cross-section (disaggregate) dynamics, I explore relationships among cross-section disaggregates for the Canadian economy. Disaggregating data by province, I show that dispersion is declining, and that the cross-section distribution is unimodal for measures of income per capita (relative to the national average). However, persistence exists in the cross-sectional ordering of relative income levels. Poor provinces remain relatively poor. Rich provinces remain relatively rich.

This persistence in rankings is consistent with findings by Green [37] in his study of economic growth in Canada.³⁹ He finds that provinces with the highest average output per capita in 1890 still had the highest in 1956, and those with the lowest continued to remain at the bottom. Differentials in income per capita

³⁷The Ontario results also indicate the robustness of the bi-modal pattern to the frequency of data.

³⁸During the period 1961-1994, Ontario accounted for 40% of Canadian GDP. The next largest contributors were Quebec (24%), British Columbia (11%) and Alberta (10%). Ontario and Quebec both have broad industry bases. In these two provinces, no one industry accounts for more than 17% of provincial GDP. By comparison, the chemicals industry contributes more than 28% of Alberta's GDP in 1989.

³⁹See also Melvin [56] and references cited therein.

between the top provinces and those at the bottom first widened (between 1890 and 1910) and then narrowed (from 1929).

I find a strikingly different pattern for data on wages, salaries and supplementary labour income (WSSLI) per employee. Ergodic distributions are bi-modal, suggesting divergence within the cross-section distribution. Persistence at the extremes is combined with a higher mass at the ends; the middle of the distribution is disappearing.

GDP per employee across manufacturing sectors also displays a bi-modal pattern. Manufacturing industries tend to remain either at the top or the bottom of the cross-section distribution. The long-run bi-modal pattern also appears in Ontario manufacturing industry data, but not for aggregations of eastern or western provinces. This suggests that movements in Ontario drive the national pattern. It reflects the wide range of manufacturing industries represented in Ontario, and the fact that it is the largest—in terms of GDP—province.

A comparison of the results for the provincial income per capita data and for the manufacturing GDP per employee data suggests, however, that it could be government transfer payments that are behind the unimodal long-run distribution. In the absence of these payments, the long-run distribution is bi-modal and the underlying pattern of economic growth in Canada is one of divergence.

For policy makers concerned with provincial inequalities, this essay provides a mixture of good and bad news. On the one hand, there is some evidence of long-run provincial convergence. On the other hand, there is no evidence to suggest that poorer provinces or industries can approximate richer provinces or industries in the short or medium run. Even in the presence of transfer payments, rankings remain unchanged within the cross-section distribution, for provincial and industrial disaggregations.⁴⁰

It appears that it is those factors which determine a region's comparative advantage (such as geographic location, climate, resource endowments, etc.) which determine the steady-state distribution of income across Canadian provinces.

⁴⁰Although it is arguable as to whether transfer payments were designed to affect rankings.

Figure 1
Distribution Dynamics

Figure 2
Log of Canadian Provincial Personal Income per Capita
(Relative to National Average) 1926-1994

Figure 3
Quantiles for Log of Provincial Personal Income per Capita (1928-1992)

Figure 4
Log of Provincial WSSLI per Employee (Relative to National Average) 1966-1994

Figure 5
Quantiles for Log of Provincial WSSL per Employee (1968-1993)

Figure 6
Log of Manufacturing Industry GDP per Employee
(Relative to Industry Sector Average) 1961:1-1994:11

Chapter 3

The Dynamic Effects of Aggregate and Disaggregate Disturbances

3.1 Introduction

The traditional approach to economic convergence, typified by Barro and Sala-i-Martin [4], Helliwell [40], and Mankiw et al [55], involves studying whether poor economies grow faster than rich economies (after controlling for various institutional details). Evidence of this “catch-up” effect is interpreted as support for the neoclassical growth model, as set out by Solow [80]. According to this view, the growth of income per capita is driven by aggregate (perhaps, technology) disturbances, which determine the long-run steady state of the model.

Another strand of the literature, exemplified by Durlauf and Johnson [29] and Quah [74] and [75], argues that by averaging out the cross-section information, important characteristics of the data disappear. These researchers argue that convergence studies should focus upon the degree of interaction between disaggregates. Using econometric methods which capture this interaction or cross-section information, they find that the conventional approach masks evidence of convergence clubs and polarisation. The pattern of cross-economy growth is consistent with the models developed by Durlauf [28], Long and Plosser [53] and Galor and Zeira [35]. In these studies, disaggregate disturbances play an important role in aggregate growth dynamics. In Durlauf [28], for example, local linkages across industries create sequential complementarities which build up over time to affect aggregate behaviour.

In Chapter 2, I find that Canadian disaggregate dynamics contain important information for explaining the pattern of convergence. These dynamics and aggregate economic growth may be the result of either aggregate or disaggregate disturbances—as predicted in the models developed by Durlauf [28] and others. However, the methodology adopted in that Chapter does not allow identification of different types of disturbance. In this Chapter, I directly identify the dynamic effects of disaggregate and aggregate disturbances using a dynamic restrictions-

based technique similar to that of Blanchard and Quah [14] and Shapiro and Watson [78].

I estimate a vector autoregression (VAR) system, assuming that an aggregate measure of economic growth and a measure of disaggregate interaction (or mobility) are affected by two types of disturbance. I interpret the first disturbance as an index that represents a common transitory regional shock; and the second as an aggregate shock. The disaggregate disturbance is identified as having no long-run impact on the level of aggregate income. An example of a disaggregate disturbance with this characteristic could be a provincial government fiscal stimulus. This might have short-run aggregate and disaggregate effects, but in the long-run it leaves aggregate economic activity unchanged. The second (aggregate) disturbance has permanent effects on aggregate income. An example of this type of disturbance might be a technology shock. The two types of disturbance are assumed to be orthogonal.

The interaction variable, which is constructed from information contained in transition probability matrices, is assumed to go to its population mean in the long run. That is, mobility is unaffected by either disturbance in the long run. If disturbances had long-term effects, it would imply that the level of interaction could drift over time. However, as shown in Chapter 2, this is not the case. The short-run dynamics of the two shocks on interaction are unrestricted.

I examine annual provincial and monthly manufacturing industry data. I find that the aggregate shock has a large positive impact on aggregate income per capita in both the short and the long run. The disaggregate disturbance initially increases income per capita; but the economy quickly adjusts. The effect of this disturbance becomes insignificant within 5 years; and accounts for little of the variation in the aggregate.

The short-run impact of the two types of disturbance on the provincial interaction measure is imprecisely estimated. The disaggregate shock provides the strongest impulse; and explains most of the variation. But its effect decays to zero within 3 years. The aggregate shock has a larger medium-term impact, which disappears within 7 years.

A similar pattern emerges with the industry data. The aggregate disturbance has a positive long-term effect on the aggregate measure, and explains most of its variation. The effect of the disaggregate disturbance is close to zero. Neither

disturbance has much of an effect on the interaction measure. Within only a few months, the impacts are indistinguishable from zero.

The results in this Chapter, confirm that disaggregate dynamics contain important information for explaining the pattern of economic growth. By ignoring disaggregate disturbances, researchers ascribe too much importance to aggregate disturbances.

The remainder of this Chapter is organised as follows. In Section 3.2, I present the methodology used to identify the disaggregate and aggregate disturbances. In Section 3.3, I discuss the economic interpretation behind the identifying restrictions. I present the results in Section 3.4; and conclude in Section 3.5.

3.2 Identification

I construct a just-identified structural VAR in the spirit of Blanchard and Quah [14]. The main steps are as follows. Assume that the two disturbances (disaggregate and aggregate) are uncorrelated at all leads and lags. The Wold Representation Theorem implies that, under weak regularity conditions, a stationary process can be represented as an invertible distributed lag of serially uncorrelated disturbances. In order to identify the underlying disturbances, it is assumed that they are linear combinations of the Wold innovations.

Let $\ln Y(t)$ and $Pr(t)$ denote the log of the aggregate variable and the interaction variable respectively. Assume that $\ln Y(t)$ and $Pr(t)$ have stochastic trends, but are not cointegrated. Let $X(t)$ be the vector $(\Delta \ln Y(t), Pr(t))'$ and $u(t)$ be a vector of disturbances $(u_1(t), u_2(t))'$. The usual tests suggest that $\Delta \ln Y(t)$ is stationary.⁴¹

These assumptions imply that $X(t)$ follows a stationary process, given by

$$(3.1) \quad X(t) = A(0)u(t) + A(1)u(t-1) + \dots = \sum_{j=0}^{\infty} A(j)u(t-j)$$

where $E[u(t)u(t)'] = I$. Equation (3.1) gives ΔY and Pr as distributed lags of the two disturbances. These disturbances are assumed to be pairwise orthogonal and the variance covariance matrix diagonal, normalized to the identity. The

⁴¹The Dickey-Fuller unit root test allows rejection of the null hypothesis of a unit root. This holds for both data sets, with or without a time trend, and up to (and including) 5 lags of the variable.

contemporaneous effect of u on Y is given by $A(0)$; subsequent lag effects are given by $A(j)$, $j \geq 1$. As X is stationary, neither disturbance has a long-run effect on ΔY or the interaction variable.

To recover the disturbances from the data, a VAR is estimated and inverted to obtain the following Wold moving average representation,

$$(3.2) \quad X(t) = e(t) + C(1)e(t-1) + \dots$$

where $e(t) = (e_1(t), e_2(t))'$, $C(0) = I$ and $E(e(t)e(t)') = \Omega$. From the proof of the Wold Theorem, this moving average representation is known to be unique.

Since the disturbances $(u_1(t), u_2(t))$ are linear combinations of the Wold innovations $(e_1(t), e_2(t))$, u_1 and u_2 can be uniquely recovered.⁴² From equations (3.1) and (3.2), the following two conditions hold: $e(t) = A(0)u(t)$ and $A(j) = C(j)A(0)$ for all j . If $A(0)$ is unique, u_1 and u_2 can be recovered. Three restrictions are imposed on $A(0)$ by,

$$(3.3) \quad E[e(t)e(t)'] = E[A(0)u(t)(A(0)u(t))'],$$

which implies that $\Omega = A(0)A(0)'$. The transitory nature of the disaggregate disturbance imposes a fourth: $\sum_{j=0}^{\infty} A(j) = A(0) \sum_{j=0}^{\infty} C(j)$.⁴³ This gives the aggregate and interaction variables as functions of current and past disaggregate and aggregate disturbances.

3.3 Interpretation

The key restriction in the identification procedure is that the disaggregate disturbance has no long-run effects on aggregate economic activity. Examples of such disturbances might include any of the following types of province or industry-specific fiscal stimuli: the introduction of a lump-sum subsidy; the construction of infrastructure (roads, hospitals, etc.); or expenditure on education or social security. Alternatively, the source of the disaggregate disturbance could be consumer optimism in a particular region.⁴⁴

⁴²If this assumption does not hold, the underlying disturbances cannot be recovered. For further details of this non-fundamentalness issue see Lippi and Reichlin [52] and Blanchard and Quah [15].

⁴³Blanchard and Quah [14] present a formal argument to show that $A(0)$ is just identified.

⁴⁴Notice that in some circumstances, an initially localised (perhaps fiscal) disturbance might have spillover effects on other areas—in which case, the long-run impact on aggregate economic activity would be non-zero. Under my identification scheme, these are treated as aggregate disturbances.

It is likely, as the above examples make clear, that there are more than two sources of disturbances, perhaps with different effects on the economy. The maintained hypothesis is as in Blanchard and Quah [14], where it is shown that if all disaggregate shocks are similar in nature, then they lie in a space that can be identified using the Blanchard-Quah decomposition.

A simple alternative approach to the identification problem is to ignore interaction completely. For example, Bayoumi and Eichengreen [7] estimate the responses to different aggregate shocks in each disaggregate in turn, ignoring the other disaggregates. Then the responses to shocks across disaggregates are compared. However, this procedure ignores the two-way interaction suggested in the models developed by Durlauf [28] and others—and allowed for in the methodology adopted in this Chapter.

3.4 Results

I examine both provincial and industrial disaggregate data. I use annual data on provincial income per capita from 1926 to 1993; and monthly data on GDP per employee for 21 manufacturing industries for the period 1974:2 to 1994:11.⁴⁵

3.4.1 Provincial data

First, I create a summary measure of provincial interaction, or “churning”, derived from transition probability matrices (see Chapter 2). This is estimated as follows.

Following Quah [68], I use a stochastic kernel equation to represent the dynamics of the cross-section distribution. The dynamics are parameterized by (M, Q) , where M is a transition probability matrix and Q is a sequence of quantiles or states. Each quantile-set pair $(Q(t), Q(t + 1))$ defines a transition probability matrix M of transitions from period t to period $t + 1$. Each element (j, k) of the transition probability matrix, M , indicates the probability of an observation starting in state j in period t , ending up in state k in period $t + 1$. Associated with each transition probability matrix is an ergodic or steady-state distribution (where it exists), which gives the probability of being in a particular quantile of the distribution in the long-run.⁴⁶

⁴⁵The industries are listed in Appendix A.

⁴⁶The ergodic distribution can be thought of as representing the unconditional probability

I use the information contained in the ergodic distribution to create the measure of cross-section interaction. Specifically, I calculate ergodic distributions for 4-state transition probability matrices, using 7-year overlapping sub-samples of provincial personal income per capita data. For example, the first sub-sample covers the period 1927 to 1933, the second covers 1928 to 1934, and so on. Each ergodic distribution consists of a set of 5 observations (one for each quantile or state boundary) for each sub-sample. I aggregate the transition information contained in each ergodic distribution as follows. Taking each set of 5 quantiles, I subtract the set for the previous year.⁴⁷ I take the modulus of these first differences, and calculate the average across the 5 moduli to create a measure of churning across each quantile set. I assign this interaction measure to the mid-year of that sub-sample.⁴⁸ In this way, I create a time-series measure of cross-section mobility for the period 1931-1988. I use this measure, a summary of the probability of moving between states within a 7-year period, as an indication of the interaction or churning among disaggregates.

Next, I estimate a bivariate VAR system, using data on aggregate personal income per capita growth and this interaction measure.⁴⁹ I allow for 2 lags, and include a constant and time trend.⁵⁰

The dynamic effects of the aggregate and disaggregate disturbances on aggregate income per capita and the provincial interaction measure, along with the two-standard deviation bands, are shown in Figures 7 to 10.⁵¹ The vertical axes show either the log of aggregate income per capita or the interaction measure; the horizontal axes show time in years.

The aggregate disturbance has a cumulative effect on per capita personal of being in a particular quantile at any given time.

⁴⁷For example, I subtract the set for the sub-sample 1927-1933 from the set for the sub-sample for 1928-1934, and so on.

⁴⁸For example, for the sub-sample 1928-1934, I assign the measure to 1931, and so on. By assigning the measure to the mid-year of the sub-sample, rather than the beginning or end, the aim is to create a measure of mobility around a certain point. However, if there are big changes in mobility at the ends of the sub-sample, it may be misleading to assign the measure to the mid-year. I assume that, given the overlapping nature of the sub-samples, it is unlikely that such occurrences will distort the main results.

⁴⁹Specifically, I use $100 \times \ln(\text{income growth})$ and $100 \times (\text{the interaction measure})$. The interaction measure is the average of the modulus of the first differences of probability measures.

⁵⁰I experimented with different lag lengths and found the results largely invariant.

⁵¹The standard deviation bands are obtained by 1,000 bootstrap replications as follows. First, I estimate the VAR and identify the two types of disturbance using the dynamic restrictions. Second, I draw with replacement from the distribution of the fitted errors. I use these to generate 1,000 sample replications. For each of these, I identify the disaggregate and aggregate disturbances, and compute the associated impulse responses.

income. After about 3 years, the peak response is almost twice the initial one. Then the effect decreases, and stabilizes to its long-run level after 7 years. The disaggregate disturbance initially increases income per capita (although the impact is much smaller than for the aggregate disturbance), but the effect falls sharply within 3 years, and disappears inside 5 years.

The response of the interaction measure to an aggregate disturbance is imprecisely estimated. The peak response is around 3 years; its long-run impact is zero, reached after about 7 years. Initially, the impact of the disaggregate disturbance on provincial interaction is quite strong (and positive), but it decays very quickly. The response dies out completely within 3 years. In this case, the impact of the disaggregate disturbance is much larger than that of the aggregate disturbance. Since the interaction measure is stationary, neither disturbance has a persistent effect on provincial interaction. Even an aggregate disturbance, such as a productivity shock, has no long-run effect.

Table 3.1 presents variance decompositions for the income per capita and provincial interaction measures. These assess the relative contribution of the different disturbances to fluctuations in the two measures at various horizons. Two standard deviation bands are shown in parentheses. The k -period ahead forecast error in aggregate income per capita is defined as the difference between the actual value of personal income per capita and its forecast from equation (3.2) as of k years earlier. The forecast error is due to both aggregate and disaggregate disturbances in the last k years. The number under aggregate income per capita at horizon k ($k = 0, 1, 2, 4, 8, 12, 20, \infty$) gives the percentage of variance of the k -year ahead forecast due to the aggregate disturbance. Impact and infinite-horizon shocks are given by the 0-year and ∞ (defined as the sample length) ahead forecasts respectively. The contribution of the disaggregate disturbance is 100 minus this number. Two-standard deviation bands are shown in parentheses.

The results suggest that the aggregate disturbance can explain nearly all the variation in aggregate income per capita. However, the disaggregate disturbance appears to explain most of the variance in the interaction measure; the contribution of the disaggregate disturbance varies from 75.7% to 80.5% over different horizons. The aggregate disturbance explains little of the variance in the interaction variable.

Percentage of Variance Due to Aggregate Disturbances		
Horizon (years)	Income per Capita	Interaction
0	0.940 (0.77, 1.00)	0.195 (0.02, 0.43)
1	0.928 (0.79, 1.00)	0.195 (0.02, 0.43)
2	0.956 (0.87, 1.00)	0.231 (0.06, 0.46)
4	0.979 (0.94, 1.00)	0.243 (0.08, 0.50)
8	0.989 (0.96, 1.00)	0.243 (0.08, 0.50)
12	0.992 (0.98, 1.00)	0.243 (0.08, 0.50)
20	0.995 (0.98, 1.00)	0.243 (0.08, 0.50)
∞	0.998 (0.99, 1.00)	0.243 (0.08, 0.50)
Percentage of Variance Due to Disaggregate Disturbances		
Horizon (years)	Income per Capita	Interaction
0	0.060 (0.00, 0.23)	0.805 (0.57, 0.97)
1	0.072 (0.00, 0.19)	0.805 (0.56, 0.96)
2	0.044 (0.00, 0.12)	0.769 (0.53, 0.93)
4	0.021 (0.00, 0.06)	0.757 (0.49, 0.92)
8	0.011 (0.00, 0.03)	0.757 (0.48, 0.92)
12	0.008 (0.00, 0.02)	0.757 (0.48, 0.92)
20	0.005 (0.00, 0.01)	0.757 (0.48, 0.92)
∞	0.002 (0.00, 0.00)	0.757 (0.48, 0.92)

Table 3.1: Variance Decomposition of Aggregate Income per Capita and Provincial Interaction

The response of the aggregate measure to an aggregate disturbance is consistent with the pattern suggested by a neoclassical growth model, where long-run changes in the level of output are driven by technology shocks. The hump-shaped response of aggregate income per capita to a disaggregate disturbance is similar to the response of output to a demand disturbance in Blanchard and Quah [14]. The forecast error variance decompositions are similar to those found in the literature (see, for example, Blanchard and Quah [14] and Cogley and Nason [20].)

3.4.2 Industrial data

I repeat the process outlined above using the monthly manufacturing industry data on GDP per employee. For the interaction measure, I estimate 4-state transition probability matrices for overlapping samples, each comprising 49 observations. I derive the summary statistic as before, as the average of the modulus of the first differences of the resulting ergodic distributions. Using the GDP per employee and interaction measures, I estimate a bivariate VAR allowing for 12 lags, and including a constant and time trend.⁵² The resulting impulse responses and associated two-standard deviation bands are shown in Figures 11 to 14.

The aggregate disturbance has a positive effect on GDP per employee. The initial response is followed by a decline to just under half its original level, recovering slightly before stabilizing at its long-run level after about 3 years. After an initial decline, the disaggregate disturbance increases GDP per employee to its starting level, but the effect is small and insignificant, and disappears within 2 years.

The effect of the aggregate disturbance on interaction is insignificantly different from zero, even in the very short term. The disaggregate disturbance has a strong (positive) initial impact, but the impact declines quickly, disappearing within 2 years.

These results are broadly similar to the results obtained using the provincial data: the aggregate disturbance has a long-run positive effect on the aggregate, but no long-run effect on the interaction measure. The disaggregate disturbance

⁵² Again, the results are largely invariant to different lag lengths and the inclusion of a time trend.

(by construction) has no long-run effect on either variable. For the industry data, it has an insignificant effect on the aggregate even in the short run. This insignificance could be because the underlying disturbances are regional rather than industry-specific.

Variance decompositions for various horizons are shown in Table 3.2. The aggregate disturbance explains nearly all the variation in aggregate GDP per employee; the disaggregate disturbance explains most of the variation in the interaction measure.

3.5 Conclusions

In this Chapter, I propose a technique to examine the impact of aggregate and disaggregate disturbances on measures of aggregate economic growth and interaction across disaggregates. The interaction variable is constructed from information contained in transition probability matrices. I use data on provincial income per capita and manufacturing industry GDP per employee. The approach is motivated by those studies (such as Durlauf [28] and others) which argue that the interaction across disaggregates contains explanatory information for the aggregate, and vice versa. If the disaggregate disturbance can be shown to have some effect on aggregate economic activity, then this strengthens the case for including disaggregate information in an explanation of aggregate growth.

I find that the aggregate disturbance has positive long-run effects on provincial income per capita and GDP per employee, and explains most of their variation. The disaggregate disturbance only matters for fluctuations in aggregate income at business cycle horizons (up to 5 years). It has little effect on GDP per employee. However, the disaggregate disturbance has important short and medium-run effects on the interaction measures. In contrast, the aggregate shock contributes little. In the long run (by construction), interaction is unaffected by either disturbance.

I argue in Chapter 2 that an explanation of economic growth and/or convergence should include disaggregate information. The results from this Chapter confirm that disaggregate disturbances contain important information for aggregate economic activity at business cycle horizons. Unfortunately, the traditional method of considering convergence concentrates on the explanatory power of

Percentage of Variance Due to Aggregate Disturbances		
Horizon (months)	GDP per Employee	Interaction
0	0.955 (0.59, 1.00)	0.049 (0.00, 0.40)
1	0.941 (0.57, 0.99)	0.056 (0.00, 0.41)
6	0.962 (0.75, 0.99)	0.084 (0.05, 0.44)
12	0.978 (0.84, 0.99)	0.090 (0.08, 0.45)
24	0.986 (0.90, 0.99)	0.093 (0.08, 0.46)
36	0.990 (0.93, 1.00)	0.094 (0.08, 0.46)
48	0.993 (0.95, 1.00)	0.094 (0.08, 0.46)
60	0.994 (0.96, 1.00)	0.094 (0.08, 0.46)
120	0.997 (0.98, 1.00)	0.094 (0.08, 0.46)
∞	0.998 (0.98, 1.00)	0.094 (0.08, 0.46)
Percentage of Variance Due to Disaggregate Disturbances		
Horizon (months)	GDP per Employee	Interaction
0	0.045 (0.00, 0.39)	0.951 (0.58, 1.00)
1	0.059 (0.01, 0.39)	0.944 (0.56, 0.99)
6	0.038 (0.01, 0.25)	0.916 (0.53, 0.95)
12	0.022 (0.01, 0.16)	0.910 (0.53, 0.92)
24	0.014 (0.01, 0.09)	0.907 (0.53, 0.92)
36	0.010 (0.00, 0.06)	0.906 (0.53, 0.92)
48	0.007 (0.00, 0.05)	0.906 (0.53, 0.92)
60	0.006 (0.00, 0.04)	0.906 (0.53, 0.92)
120	0.003 (0.00, 0.02)	0.906 (0.53, 0.92)
∞	0.002 (0.00, 0.02)	0.906 (0.53, 0.92)

Table 3.2: Variance Decomposition of Aggregate GDP per Employee and Manufacturing Industry Interaction

aggregate shocks. Under the neoclassical growth model, long-run changes in output are driven by technology shocks. To gain a more complete picture of convergence in the short to medium run however, the researcher should include disaggregate information.

Figure 7
Impulse Response: Aggregate Disturbance on Aggregate Income per Capita

Figure 8
Impulse Response: Disaggregate Disturbance on Aggregate Income per Capita

Figure 9
Impulse Response: Aggregate Disturbance on Interaction

Figure 10
Impulse Response: Disaggregate Disturbance on Interaction

Figure 11
Impulse Response: Aggregate Disturbance on Aggregate GDP per Employee

Figure 12
Impulse Response: Disaggregate Disturbance on Aggregate GDP per Employee

Figure 13
Impulse Response: Aggregate Disturbance on Interaction Measure

Figure 14
Impulse Response: Disaggregate Disturbance on Interaction Measure

Chapter 4

Testing For Canadian Unit Roots: A Panel Data Approach

4.1 Introduction

To apply standard unit root tests to panel data, a researcher must aggregate across cross-sections, discarding the cross-sectional or disaggregate information. Recently, more powerful unit root tests have been developed for panel data, which allow cross-sectional information to be included.⁵³ Levin and Lin (LL) [51] and Quah [75] show that implementing a unit root test which incorporates both cross-section and time-series information can dramatically improve power.

In this Chapter, I present the results from a variety of unit root tests on panel data. I compare the results of the standard Dickey-Fuller method with those from three tests designed specifically for panels. Unlike the Dickey-Fuller test, the panel data unit root tests have limiting normal distributions.

The first test takes the standard approach to unit root tests, but, to allow for fixed effects, the initial observation for each individual time-series is subtracted from each subsequent observation in that series. The unit root hypothesis is tested by computing a standard t-ratio using data that is averaged across the cross-section. I refer to this as the “panel Dickey-Fuller” test (Breitung and Meyer (BM) [16]).

The second test requires the data to be adjusted by subtracting the cross-section mean. The test statistic is the average of the individual coefficients on the lagged dependent variable. I refer to this as the “ δ -bar” test (LL [51]). I extend the LL [51] methodology to allow for potentially more complex common effects across individual series. Using a dynamic index model developed by Quah and Sargent [76], I identify multiple common factors and subtract the resulting indeces (in place of the cross-section average) from the data.

In the third approach, the “ t -bar” test of Im, Pesaran and Shin (IPS) [42], the initial observation for each individual time-series is subtracted from each

⁵³The importance of including cross-sectional information is illustrated in the context of economic convergence in Chapter 2. There, I show that ignoring cross-section information can mask underlying patterns of mobility and shape.

subsequent observation in that series—in a similar manner to the panel Dickey-Fuller test. The test statistic is a (small sample adjusted) average of the Dickey-Fuller test statistics for the individual series.

I examine Canadian annual data on gross provincial product (GPP) for the period 1961-1990 and monthly data on gross domestic product (GDP) per employee for 19 manufacturing industries over the period 1974:2 to 1994:11. When I apply the standard Dickey-Fuller test, to data averaged across the cross-section, I cannot reject the null hypothesis of a unit root in either the provincial or industrial data. In contrast, none of the three panel unit root tests suggest that the data are characterised by a unit root.

The presence of a unit root has important implications for macroeconomic studies. If data contain a unit root, shocks will have permanent effects; where there is no unit root, shocks will have only temporary effects.⁵⁴ The results from this Chapter suggest that there are no permanent shocks in Canadian annual GPP and monthly manufacturing industry GDP per employee data.

In a seminal paper, Granger [36] argues that the aggregation of a multivariate time series process can produce a univariate process that has fundamentally different properties from the multivariate process. The results reported here are consistent with Granger's finding: aggregating the data (in order to use an existing test) can radically alter the dynamics of the process.

The rest of this Chapter is arranged as follows. In Section 4.2, I review the related literature. In Section 4.3, I outline the theoretical methodology for each of the three panel data unit root tests used in this Chapter. I present the results in Section 4.4, and conclude in Section 4.5.

4.2 Unit Roots and Panel Data

A linear process contains a unit root if one of the roots or eigenvalues of the autoregressive polynomial is unity and all other eigenvalues are inside the unit circle. Hypothesis tests concerning the coefficients of non-stationary variables cannot be conducted using tests based on the asymptotic t - and F -distributions. The distributions of the t - and F -statistics are non-standard.

Dickey and Fuller [25] and [26] test for unit roots by deriving a non-standard

⁵⁴See, for example, Nelson and Plosser [57].

distribution using Monte Carlo simulations. They consider three different models:

$$(4.1) \Delta y_t = \delta y_{t-1} + \epsilon_t$$

$$(4.2) \Delta y_t = \alpha + \delta y_{t-1} + \epsilon_t$$

$$(4.3) \Delta y_t = \alpha + \delta y_{t-1} + \gamma t + \epsilon_t.$$

The first is a pure random walk model, the second adds an intercept, and the third includes both an intercept and linear time trend. If $\delta = 0$, the $\{y_t\}$ sequence contains a unit root. The test involves estimating one (or more) of the equations above using OLS to obtain the estimated value of δ and its associated standard error. Comparing the resulting t -statistic with the appropriate critical value (reported in the Dickey-Fuller tables) allows the researcher to determine whether to reject the null hypothesis that $\delta = 0$. The critical values depend on the form of regression and sample size.

The same critical values are used to test for the presence of a unit root in higher order equations. An augmented Dickey-Fuller (ADF) test is used when ϵ_t is serially correlated. The empirical model contains a number of lags of the dependent variable,

$$(4.4) \Delta y_t = \alpha + \delta y_{t-1} + \sum_{L=1}^p \theta_L \Delta y_{t-L} + \epsilon_t$$

where p is the order of autoregression.

In Section 4.4.3 below, I estimate equation (4.4) (and a similar equation including a time trend) using Canadian annual data on GPP for 1961-1990, and monthly data on GDP per employee for 19 manufacturing industries over the period 1974:2 to 1994:11. In order to be able to use the Dickey-Fuller test, I average the data across the cross-section. I find that I cannot reject the null hypothesis of a unit root for either of the data series.

These results are consistent with those of Otto and Wirjanto [58], for example, who test quarterly Canadian macro time series (1955:1-1988:4) for seasonal and non-seasonal unit roots. Their results suggest that most of the series, including GDP, contain a unit root. Banerjee, Lumsdaine and Stock [2], on the other hand, compare the null hypothesis of a unit root with stationarity about

a broken trend. Using quarterly Canadian GNP data (1948:1-1989:2), they reject the null hypothesis in favour of the trend-shift alternative, with a break in 1981:3.⁵⁵ Their interpretation is that the recession of the early 1980s led to a permanent downward shift in the trend growth rate. After the recovery, output is stationary around its original growth path.

However, averaging across the cross-section loses cross-section or disaggregate information. LL [51] show, for example, that, in the absence of individual-specific effects, the power of the standard Dickey-Fuller test is low for short time series ($T \leq 50$). Allowing for both individual-specific intercepts and time trends, the Dickey-Fuller test has very low power even for longer time series ($T \leq 100$).

In a theoretical paper, Quah [71] examines the characteristics of the time-series unit root coefficient estimator in panel data with simultaneously extensive cross-section and time-series dimensions. He finds that the asymptotic distribution of the estimator is neither Dickey-Fuller (as for standard time series analysis) nor normal and asymptotically unbiased (as for standard panel data analysis). It is consistent and asymptotically normal, but has a non-vanishing bias in its asymptotic distribution.

Three recent papers propose alternative unit root tests, designed specifically for panel data, which evaluate the null hypothesis that each individual series has a unit root against the hypothesis that all series are stationary. Unlike the standard Dickey-Fuller test, each of these panel data tests has a limiting normal distribution.

BM [16] apply their panel Dickey-Fuller test to German wage data (1972-1987). The initial observation for each individual time series is subtracted from each subsequent observation in that series before a standard t -test is used to test the unit root hypothesis. To obtain the limiting distribution of the test statistic, BM [16] let the number of cross-sectional units tend to infinity. In this case, the null distribution becomes asymptotically normal rather than non-standard, as when the number of time periods approaches infinity. Allowing for both a linear time trend and random time effects, the unit root hypothesis cannot be rejected for firm and industry-level wages.

LL's [51] δ -bar test statistic is an average of the individual coefficients on

⁵⁵The result is sensitive to the number of autoregressive lags used.

the lagged dependent variable.⁵⁶ Monte Carlo simulations indicate that the normal distribution provides a good approximation to the empirical distribution of the δ -bar test statistic, even for relatively small panels.⁵⁷ The null hypothesis imposes a cross-equation restriction on the first-order partial autocorrelation coefficients, enabling the δ -bar test procedures to yield much higher power (against stationary alternatives) than performing a separate unit root test for each individual. Simulations indicate that this increase in power holds when pooling the time-series data for even a small number of cross-sectional units ($N \geq 10$). With individual-specific intercepts and trends, the δ -bar test is powerful with relatively moderate-sized panels ($N = 10$ and $T = 50$ or $N = 25$ and $T = 25$). The δ -bar test procedures are most appropriate for data with 10 – 250 individuals, and 25 – 250 time series observations per individual.⁵⁸

In an application of the δ -bar test, MacDonald [54] examines a group of OECD real exchange rates over the recent floating experience (1973-1992). Regardless of whether or not a time trend is included, he rejects the null hypothesis of a unit root. Using standard Dickey-Fuller tests, on the other hand, a unit root cannot be rejected.

The t -bar test is a small-sample adjustment to the average of the Dickey-Fuller test statistics for each individual series (calculated after first adjusting the data by subtracting the initial observation from each time-series observation). IPS [42] show that it is valid with heterogeneity across groups and in the presence of residual serial correlation across time periods. Under the null hypothesis of a unit root, the t -bar statistic has a standard normal distribution for N (the number of cross-sectional units) and T (the number of time periods) sufficiently large and $N/T \rightarrow 0$, and diverges to $-\infty$ under the alternative hypothesis of stationarity. Using Monte Carlo simulations, IPS [42] show that the t -bar statistic is superior in small samples to the δ -bar test: in general, the δ -bar test does not converge to the standard normal distribution, even as $N \rightarrow \infty$, $T \rightarrow \infty$ and $N/T \rightarrow 0$.

Lee, Pesaran and Smith [48] present an application of the t -bar test using

⁵⁶The data are first adjusted by subtracting the cross-section average from each observation.

⁵⁷LL [51] show that the δ -bar statistic has a limiting normal distribution as both the cross-section and time-series dimensions of the panel grow large. The standard error of the estimated empirical power was less than 0.01 in all cases (with 10,000 replications).

⁵⁸Existing tests are preferable for samples with either longer time-series or more cross-sections.

output per capita data for 102 countries over the period 1960-1989. They reject the existence of a unit root. Using a standard Dickey-Fuller test, they find that the unit root hypothesis is only rejected for 8 out of the 102 countries.

In the next Section, I outline in more detail the methods used to calculate the unit root tests: the panel Dickey-Fuller test developed by BM [16], the δ -bar test developed by LL [51] (and my extension), and IPS's [42] t -bar test.

4.3 Methods

4.3.1 The panel Dickey-Fuller test

Assume that the series y_{it} , has a finite order autoregressive representation, where the autoregressive parameters do not vary over time and individuals in the cross-section. Consider the $AR(1)$ model given by,

$$(4.5) \quad y_{it} = \delta y_{i,t-1} + (1 - \delta)\mu_i + \epsilon_{it}$$

for $t = 1, \dots, T$ and $i = 1, \dots, N$, with $E(y_{it}) = \mu_i$. The ϵ_{it} are uncorrelated errors distributed $N(0, \sigma^2)$. Under the alternative hypothesis of no unit root, $|\delta| < 1$, the OLS estimate $\hat{\delta}$ is biased against $\delta = 1$ (leading to a loss of power).

To overcome this problem, subtract the initial observation, y_{i0} , from both sides of equation (4.5) to get,

$$(4.6) \quad x_{it} = \delta x_{i,t-1} + \eta_{it}$$

for $t = 2, \dots, T$, $i = 1, \dots, N$, where $x_{it} = y_{it} - y_{i0}$, $x_{i,t-1} = y_{i,t-1} - y_{i0}$ and $\eta_{it} = \epsilon_{it} + (1 - \delta)(y_{i0} - \mu_i)$. BM [16] show that under the null hypothesis of a unit root the asymptotic bias disappears. A conventional t -test on the OLS estimate $\tilde{\delta}$ is used to test $H_0 : \tilde{\delta} = 1$.

To introduce a linear time trend,⁵⁹ the model is modified to give,

$$(4.7) \quad y_{it} = \delta y_{i,t-1} + \gamma^* t + \mu_i^* + \epsilon_{it}$$

where $\gamma^* = (1 - \delta)\gamma$ and $\mu_i^* = (1 - \delta)\mu_i + \delta\gamma$.⁶⁰ Ignoring individual effects leads to biased estimates of δ and γ^* under the alternative hypothesis of stationarity when using OLS. As for the basic model, subtract the initial observation from both sides of equation (4.7) to get,

⁵⁹Of the form $E(y_{it}) = \mu_i + \gamma t$.

⁶⁰This is derived from $(1 - \delta L)(\mu_i + \gamma t) = (1 - \delta)(\mu_i + \gamma t) + \delta\gamma$.

$$(4.8) \quad x_{it} = \delta x_{i,t-1} + \gamma^* t + \epsilon_{it} - (1 - \delta)(y_{i0} - \mu_i).$$

where x_{it} and $x_{i,t-1}$ are defined as before. Under the null hypothesis of a unit root, the last term vanishes and OLS estimates of δ and γ^* are consistent. The distribution of the t -statistic for $\delta = 1$ does not depend on individual effects. The limiting distribution for $N \rightarrow \infty$ does not change if a time trend is included. (Although the limiting distribution for $T \rightarrow \infty$ is affected by including a time trend.)

For wide variation in individual effects, the performance of the panel Dickey-Fuller test is superior to the standard ADF approach (when cross-sectional information is important). However, it does not permit the coefficient on $y_{i,t-1}$ or the lag order to vary across individuals: all individual effects are assumed to be incorporated by subtracting the initial time-series value from all subsequent observations.

4.3.2 The δ -bar test

Let a stochastic process be represented by $\{y_{it}\}$ for individuals $i = 1, \dots, N$ over time periods $t = 1, \dots, T$. Assume (as for the panel Dickey-Fuller test) that all individuals in the panel have identical first-order partial autocorrelation, but all other parameters of the error process are unrestricted across individuals.

LL [51] present δ -bar test procedures for three different models: when the series y_{it} (i) has zero mean for every individual in the panel (equation (4.9)); (ii) has an individual-specific mean (equation (4.10)); and (iii) has an individual-specific mean and an individual-specific time trend (equation (4.11)).

$$(4.9) \quad \Delta y_{it} = \delta_i y_{it-1} + \zeta_{it}$$

$$(4.10) \quad \Delta y_{it} = \alpha_i + \delta_i y_{it-1} + \zeta_{it}$$

$$(4.11) \quad \Delta y_{it} = \alpha_i + \gamma_i t + \delta_i y_{it-1} + \zeta_{it}$$

where $-2 < \delta_i \leq 0$ for all $i = 1, \dots, N$. The disturbance ζ_{it} is distributed independently across individuals and follows a stationary invertible ARMA process for each individual.⁶¹ Under the null hypothesis, each individual time series

⁶¹ Given by,

$$\zeta_{it} = \sum_{j=0}^{\infty} \beta_{ij} \zeta_{it-j} + \epsilon_{it}.$$

has a unit root.⁶² Under the alternative hypothesis, the process $\{y_{it}\}$ is trend-stationary for each individual in the panel.⁶³ I outline the test procedures for Model 1.

The δ -bar test requires the data to be generated independently across individuals. This assumption can be relaxed to allow for a limited amount of time-dependence via time-specific aggregate effects. As a first step, LL [51] remove the influence of aggregate effects by subtracting cross-section averages from the data. The next step calculates orthogonalized first differences and lagged levels for each individual in the cross-section by estimating the following equations (using the adjusted data),

$$(4.12) \quad \hat{e}_{it} = \Delta x_{it} - \sum_{L=1}^{p_i} \hat{\theta}_{1iL} \Delta x_{it-L}$$

$$(4.13) \quad \hat{v}_{it} = x_{it-1} - \sum_{L=1}^{p_i} \hat{\theta}_{2iL} \Delta x_{it-L}.$$

where $x_{it} = y_{it} - \bar{y}_{it}$ and $\bar{y}_{it} = \frac{1}{N} \sum_{i=1}^N y_{it}$. Lag orders (p_i) can vary across individuals.⁶⁴ Then, regress the orthogonalized innovations on the orthogonalized lagged level,

$$(4.14) \quad \hat{e}_{it} = \delta_i \hat{v}_{it-1} + \epsilon_{it}.$$

To control for heterogeneity across individuals, $\hat{e}_{it}$ and $\hat{v}_{it-1}$ are normalized by dividing by the estimated residual standard error from equation (4.14),⁶⁵ to form $\tilde{e}_{it}$ and $\tilde{v}_{it-1}$, where,

$$(4.15) \quad \tilde{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{\epsilon i}}$$

⁶²Under Model 1, $\delta_i = 0$ for all $i = 1, \dots, N$; under Model 2, $\delta_i = 0$ and α_i is zero, for all $i = 1, \dots, N$; under Model 3, $\delta_i = 0$ and γ_i is zero, for all $i = 1, \dots, N$.

⁶³Under Model 1, $\delta_i < 0$ for all $i, \dots, N$; under Model 2, $\delta_i < 0$ and α_i is non-zero for all $i, \dots, N$; under Model 3, $\delta_i < 0$ and γ_i is non-zero for all $i, \dots, N$.

⁶⁴The appropriate lag order p_i is selected as follows: for a given sample length T , a maximum lag order p_{max} is chosen, and then the t -statistics on the θ_{it} are used to determine if a smaller lag can be used.

⁶⁵The regression standard error is derived from,

$$\hat{\sigma}_{\epsilon i}^2 = \frac{1}{T - p_i - 1} \sum_{t=p_i+2}^T (\hat{e}_{it} - \delta_i \hat{v}_{it-1})^2.$$

and,

$$(4.16) \quad \tilde{v}_{it-1} = \frac{\hat{v}_{it-1}}{\hat{\sigma}_{ei}}.$$

Next, calculate the long run standard deviation ($\hat{\sigma}_{xi}$) as the variation of Δx_{it} at zero frequency.⁶⁶ If the data contain a time trend (Model 3), the trend is removed before σ_{xi} is calculated.⁶⁷

In the next step, the ratios of long-run to short-run standard deviations for each individual ($\hat{s}_i = \frac{\hat{\sigma}_{xi}}{\hat{\sigma}_{ei}}$) and the average ratio for the panel ($\hat{S} = \frac{1}{N} \sum_{i=1}^N \hat{s}_i$) are calculated. The average ratio is used to adjust the mean of the unit root test statistic.

Finally, estimate the following equation,

$$(4.17) \quad \tilde{e}_{it} = \delta_i \tilde{v}_{it-1} + \tilde{\epsilon}_{it}.$$

This enables several statistics to be calculated: the least squares estimate, $\hat{\delta}$; the standard error of the regression, $\hat{\sigma}_\epsilon$; the reported standard error of $\hat{\delta}$, $RSE(\hat{\delta})$; and the regression t -statistic for testing $H_0 : \delta = 0$, t_δ .⁶⁸ The δ -bar test statistic

⁶⁶The long run standard deviation is given by,

$$\hat{\sigma}_{xi}^2 = \frac{1}{T-1} \sum_{t=2}^T \Delta x_{it}^2 + 2 \sum_{L=1}^K w_{KL} \left(\frac{1}{T-1} \sum_{t=2+L}^T \Delta x_{it} \Delta x_{it-L} \right)$$

where $w_{KL} = L/(K+1)$.

⁶⁷ $(\Delta x_{it} - \Delta \bar{x}_{it})$, where $\Delta \bar{x}_{it}$ is the average value of Δx_{it} for individual i , is substituted for Δx_{it} in the equation given in the previous footnote.

⁶⁸These statistics are defined as,

$$\hat{\delta} = \frac{\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{it-1} \tilde{e}_{it}}{\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{it-1}^2},$$

$$\hat{\sigma}_\epsilon = \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=2+p_i}^T (\tilde{e}_{it} - \hat{\delta} \tilde{v}_{it-1})^2 \right)^{1/2},$$

$$RSE(\hat{\delta}) = \hat{\sigma}_\epsilon \left(\sum_{i=1}^N \sum_{t=2+p_i}^T \tilde{v}_{it-1}^2 \right)^{-1/2},$$

and

$$t_\delta = \frac{\hat{\delta}}{RSE(\hat{\delta})}.$$

t_δ^* is then given by,

$$(4.18) \quad t_\delta^* = \frac{t_\delta - N \tilde{T} \hat{S} \hat{\sigma}_\epsilon^{-2} RSE(\hat{\delta}) \mu_{\tilde{T}}^*}{\sigma_{\tilde{T}}^*}$$

where $\tilde{T} = (T - \bar{p} - 1)$ is the average number of observations per individual in the panel and $\bar{p} = \frac{1}{N} \sum_{i=1}^N p_i$. The mean and standard deviation adjustments, $\mu_{\tilde{T}}^*$ and $\sigma_{\tilde{T}}^*$ respectively, are given in LL [51].⁶⁹

LL [51] show that the δ -bar test statistic is distributed $N(0, 1)$ in the limit, so the critical values of the standard normal distribution are used to test the null hypothesis that $\delta_i = 0$ for all $i = 1, \dots, N$.

LL [51] implicitly assume that there is a single aggregate common factor that has an identical impact on all individuals. I allow for more complex contemporaneous correlation in the δ -bar test by, at the first step, subtracting index values (rather than the cross-section average) from the data. The indices are calculated using the methods of Quah and Sargent (QS) [76].⁷⁰

The QS approach does not specify in advance any particular pattern of co-movements; it uses the orthogonality properties of permanent and common components to characterise the model. Each y_i , $i = 1, \dots, N$ is assumed to be affected by J common factors u_j , $j = 1, \dots, J$ (common across all i), and an idiosyncratic disturbance z_i , $i = 1, \dots, N$ (specific to each i and orthogonal to all other i). The common factors are assumed to impact on y_i with M_a lags. The first differences of each u_j are pairwise orthogonal, and have a finite autoregressive representation (of order M_g). The idiosyncratic disturbance, z_i , is a finite order autoregression (of order M_b).

Translating the model into state space form,⁷¹ QS [76] exploit the fact that the common factors have dimension $O(J)$ ($J \leq N$), which is independent of $O(N)$. They use standard Kalman smoother calculations to obtain conditional

⁶⁹The adjustments vary with each model. In calculating the adjustments, LL [51] set $\bar{K} = 3.21T^{1/3}$ and $p_i = 0$ for each $i = 1, \dots, N$. Gaussian random numbers with unit variance are used to generate 250 independent random walks of length $\tilde{T} + 1$. These data were used to construct the sample statistics $\hat{S}$, $\hat{\delta}$, $\hat{\sigma}_\epsilon$, $RSE(\hat{\delta})$ and t_δ . Based on 25,000 replications, the adjustment factors $\mu_{\tilde{T}}^*$ and $\sigma_{\tilde{T}}^*$ are estimated as the mean values of $\left[\frac{t_\delta}{N \tilde{T} \hat{S} \hat{\sigma}_\epsilon^{-2} RSE(\hat{\delta})} \right]$ and the standard deviation of $[t_\delta - N \tilde{T} \hat{S} \hat{\sigma}_\epsilon^{-2} RSE(\hat{\delta})]$ respectively.

⁷⁰Appendix F outlines the Quah and Sargent [76] methodology in more detail.

⁷¹State space representation is a means of summarizing finite processes. More details of the translation process are in QS [76].

expectations.⁷² Iteration on this scheme is the EM algorithm.⁷³

4.3.3 The t -bar test

Consider, as before, a sample of N cross-sectional units observed over T time periods. For each cross-sectional unit and each time period, assume that the stochastic process y_{it} is generated by the first-order autoregressive process,

$$(4.19) \quad y_{it} = (1 - \phi_i)\mu_i + \phi_i y_{i,t-1} + \epsilon_{it}$$

where $\mu_i = \mu + \eta_i$ and $\epsilon_{it} = A_i(L)e_{it}$ with $A_i(L) = a_{i0} + a_{i1}L + a_{i2}L^2 + \dots$, $a_{i0} = 1$, $\sum_{j=0}^{\infty} a_{ij}^2 < \infty$, $i = 1, 2, \dots, N$ and L is a lag operator. Assume that the e_{it} are independently distributed across both groups and time periods with zero mean and finite variance. As for the panel Dickey-Fuller test, subtract the initial observations from each side of equation (4.19) to get,

$$(4.20) \quad x_{it} = \alpha_i + \theta_i x_{i,t-1} + \epsilon_{it}$$

where $x_{it} = y_{it} - y_{i0}$, $x_{i,t-1} = y_{i,t-1} - y_{i0}$ and $\alpha_i = (1 - \phi_i)(\mu_i - y_{i0})$. Taking first differences,

$$(4.21) \quad \Delta x_{it} = \alpha_i + \delta_i x_{i,t-1} + \epsilon_{it}$$

where $\delta_i = \theta_i - 1$.

Combine equation (4.21) with the expression for the error term, $\epsilon_{it} = A_i(L)e_{it}$. Since $A_i(L)$ is invertible, and assuming a finite dimensional specification for ϵ_{it} ,⁷⁴ produces the following augmented Dickey-Fuller equation,

$$(4.22) \quad \Delta x_{it} = \alpha_i + \delta_i x_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta x_{i,t-L} + v_{it}$$

⁷²The Kalman filter is a tool used to estimate the state vector in an optimal way. Kalman smoothing is an inference about the value of the state vector based on the full set of data collected, e.g. estimate for 1980 using data from a full sample for the period 1961 to 1990.

⁷³The EM algorithm is a method for maximizing a likelihood function in the presence of missing observations. It consists of two steps, an estimation and a maximization step, which are iterated to convergence. The maximization step calculates the maximum likelihood estimates of all unknown parameters conditional on a full data set. The estimation step constructs estimates of the sufficient statistics of the problem conditional on the observed data and the parameters. Missing observations are estimated based on the parameter values at one step of the iteration and then the likelihood function is maximized assuming that this is the full observable data set in the other (see, for example, Watson and Engle [81]).

⁷⁴Approximate ϵ_{it} by, for each $i = 1, \dots, N$, $\epsilon_{it} = \rho_{i1}\epsilon_{i,t-1} + \rho_{i2}\epsilon_{i,t-2} + \dots + \rho_{ip_i}\epsilon_{i,t-p_i} + v_{it}$ for $t = 1, 2, \dots, T$.

where the lag length p_i is allowed to vary across individuals. The null hypothesis of a unit root $H_0 : \delta_i = 0$ for all i , is tested against the stationary alternative $H_1 : \delta_i < 0$ for all i .

Where the stochastic process generating y_{it} contains a linear trend, equation (4.22) is modified to,

$$\Delta x_{it} = \alpha_i + \gamma_i t + \delta_i x_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta x_{i,t-L} + v_{it}$$

where the v_{it} are assumed to be distributed independently across both i and t .

IPS [42] base the t -bar test on the group mean of the t -statistics for testing $\delta_i = 0$,

$$(4.23) \quad \bar{t}_{NT}(p, \rho) = \frac{1}{N} \sum_{i=1}^N t_{iT}(p_i, \rho_i)$$

where p_i is (unknown but) estimated as $T^{1/3}$, $p = (p_1, p_2, \dots, p_N)'$, $\rho_i = (\rho_{i1}, \rho_{i2}, \dots, \rho_{ip_i})'$ and $\rho = (\rho_1, \rho_2, \dots, \rho_N)$. Assuming that $N/T \rightarrow 0$ as $N \rightarrow \infty$ and $T \rightarrow \infty$,⁷⁵ the standardized t -bar statistic is given by,

$$(4.24) \quad \bar{z}_{NT} = \frac{N^{1/2}[\bar{t}_{NT} - a_{NT}]}{b_{NT}^{1/2}}$$

which is asymptotically distributed $N(0, 1)$. The mean (a_{NT}) and variance (b_{NT}) of the standard unit root distribution are given by,

$$a_{NT} = \frac{1}{N} \sum_{i=1}^N E[t_T(p_i, 0)]$$

and

$$b_{NT} = \frac{1}{N} \sum_{i=1}^N V[t_T(p_i, 0)],$$

respectively. IPS [42] present tabulated values of $E[t_T(p_i, 0)]$ and $V[t_T(p_i, 0)]$ for different values of T and $p = 0, 1, \dots, 12$, for cases when the underlying ADF regressions for each group of individuals are estimated both with and without a linear trend.⁷⁶ For T large enough, these tabulated values are valid for normal and non-normal errors.

⁷⁵These assumptions are necessary because in practice $\bar{z}_{NT}$ depends on the parameters p_i and is therefore not an operational statistic.

⁷⁶The mean ($E[t_T(p_i, 0)]$) and variance ($V[t_T(p_i, 0)]$) are computed using stochastic simulations with 50,000 replications. For the model with no time trend, $t_T(p, 0)$ is the t -statistic

4.4 Results

In this Section, I present the results from applying each of the three panel data unit root tests to Canadian data, and comparing these with the outcome when using a standard Dickey-Fuller test which is applied to data averaged across the cross-section. I use two different sets of data. First, I apply the tests to annual CANSIM data on gross provincial product (GPP) for the ten Canadian provinces for the period 1961-1990. These data are deflated by a national price index.⁷⁷ Then I use monthly data on GDP per employee for 19 different manufacturing industries across Canada over the period 1974:2 to 1994:11.⁷⁸

4.4.1 Provincial data

Panel Dickey-Fuller test

After adjusting the data by subtracting the initial value from each observation, I estimate equation (4.6) using OLS. The resulting equation is given below,

$$x_{it} = 0.08273 + 0.9743x_{i,t-1} \\ (0.01118) \quad (0.00985)$$

where standard errors are shown in parentheses. The t -statistic for testing $\delta = 1$ is -2.61 . The 95% critical value for 298 degrees of freedom is 1.645. Hence, using the panel Dickey-Fuller test, I can infer that the provincial data do not contain a unit root.

of $\hat{\delta}$ in the ADF regression given by,

$$\Delta x_t = \hat{\alpha} + \hat{\delta}x_{t-1} + \sum_{L=1}^p \hat{\theta}_L \Delta x_{t-L} + error.$$

Where a time trend is included, the following equation is used,

$$\Delta x_t = \hat{\alpha} + \hat{\delta}x_{t-1} + \hat{\gamma}t + \sum_{L=1}^p \hat{\theta}_L \Delta x_{t-L} + error.$$

The underlying data generating process is $\Delta x_t = \epsilon_t$, where $\epsilon_t \sim N(0, 1)$, $t = 1, \dots, T$.

⁷⁷See Coulombe and Lee [21].

⁷⁸I use data for the following industries: food, beverage, tobacco, rubber, leather and allied products, primary textiles and textile products, clothing, wood, furniture and fixtures, paper, primary metals, fabricated metals, machinery, transport equipment, electrical and electronics, non-metallic mineral products, refined petroleum, chemicals and other.

	Model 2	Model 3
$\hat{\delta}$	-0.152120	-0.226170
$\hat{\sigma}_\epsilon$	0.990481	0.971396
$RSE(\hat{\delta})$	0.033338	0.043403
t_δ	-4.562972	-5.210919
t_δ^*	-4.177988	3.956520

Table 4.1: δ -Bar Test Statistics for Provincial Data

Including a time trend does not change the result. I estimate equation (4.8) and find,

$$x_{it} = 0.07540 + 0.01185t + 0.94525x_{i,t-1}.$$

(0.01141) (0.00446) (0.01465)

In this case, the t -statistic for the hypothesis $\delta = 1$ equals -3.74 . Once again, this value exceeds the 95% critical value of 1.645. The data do not support the null hypothesis of a unit root.

δ -bar test: base case

I assume initially that the data are generated by Model 2, i.e. the data may have an individual specific mean, but do not contain a time trend. I subtract the national average for each time period from each observation. Then I regress Δx_{it} and x_{it-1} on the first two lags of the first differences and a constant. This produces the residuals $\hat{e}_{it}$ and $\hat{v}_{it-1}$. Using the weights and lag truncation parameters recommended by LL [51], I derive the average standard deviation ratio ($\hat{S}$). Finally, estimating equation (4.17) and computing t_δ^* as in equation (4.18) produces the test statistics presented in Table 4.1.

Since $t_\delta^* \sim N(0, 1)$, I use the critical values of the standard normal distribution to test the null hypothesis that $\delta_i = 0$ for all $i = 1, \dots, N$. The absolute value of the δ -bar test statistic is 4.18. This clearly exceeds the 95% critical value of 1.96; the data reject the null hypothesis of a unit root.

Under Model 3, the series $\{y_{it}\}$ may have an individual specific mean and a time trend. The results from estimation of this model are also given in Table 4.1. The δ -bar test statistic of 3.96 clearly exceeds the 95% critical value of 1.96. The data continue to reject the null hypothesis of a unit root.

	Model 2(1)	Model 2(2)	Model 3(1)	Model 3(2)
$\hat{\delta}$	-0.4985409	-0.5435792	-0.550403	-0.594886
$\hat{\sigma}_\epsilon$	1.019640	1.016973	1.022312	1.020333
$RSE(\hat{\delta})$	0.0653325	0.0670925	0.058905	0.060517
t_δ	-7.630826	-8.101936	-9.343975	-9.830071
t_δ^*	-92.16263	-96.18277	-96.76558	-100.8222

Table 4.2: δ -Bar Test Statistics (with Index Adjustment) for Provincial Data

δ -bar test: index adjustment

I now repeat the δ -bar test, adjusting the data by subtracting an index at the first step, rather than the cross-section average. Using the QS [76] dynamic index model, I assume that $M_a = 1$, $M_g = 1$ and $M_b = 2$ for both a one- and a two-index model. The one-index model converges after 28 iterations; the two-index converges after 14 iterations.⁷⁹

The results for Model 2 (constant but no time trend) and Model 3 (constant plus time trend) are presented in Table 4.2.⁸⁰ Adjusting the data by subtracting either a single index or the sum of two indices in the two-index representation, the data continue to reject the null hypothesis that $\delta_i = 0$ for both models (the 95% critical value is 1.96). The δ -bar test statistics vary little with the number of indices or the model used.

t -bar test

I adjust the data by subtracting the initial value from each observation, and estimate equation (4.22) (constant but no time trend) for each province. For most provinces, $p_i = 0$; for Quebec, $p_i = 1$; and for PEI, Nova Scotia and Saskatchewan, $p_i = 4$. The t -statistics for testing $\delta_i = 0$ are shown in Table 4.3.

Values for the mean (a_{NT}) and variance (b_{NT}) of the standard unit root distribution are calculated as -1.48 and 0.83 respectively, giving a value for the standardized t -bar statistic of 3.14 . The 95% critical value for the normal

⁷⁹Appendix F includes charts comparing the explanatory power of two observable measures—aggregate employment and national GNP—with one- and two-index representations for the provincial data. The one-index (one common factor) model provides little explanatory power over the observable measures, but the two-index (two common factors) model appears to provide a better description of the underlying common comovements.

⁸⁰Model $i(j)$ refers to the estimation of Model i using a j -index adjustment.

Province	<i>t</i> -Statistic: No Time Trend	<i>t</i> -Statistic: Time Trend
Newfoundland	-0.39453	-2.9795
PEI	0.63051	-2.6298
Nova Scotia	1.0428	-1.8176
New Brunswick	-0.37746	-2.8174
Quebec	-0.76778	-3.4542
Ontario	-0.45904	-2.1115
Manitoba	-1.2389	-1.1032
Saskatchewan	-1.7646	-1.7158
Alberta	-1.3243	-0.34315
British Columbia	-1.0687	-2.0090

Table 4.3: *t*-Statistics (*t*-Bar Test) for Provincial Data

distribution is 1.96. Therefore, the null hypothesis of a unit root continues to be rejected.

Table 4.3 also includes the *t*-statistics for testing $\delta_i = 0$ when a time trend is included in the initial regression. The values for p_i are the same as for Model 2. The mean and variance statistics are calculated as $a_{NT} = -2.11$ and $b_{NT} = 0.72$ respectively, and the standardized *t*-bar statistic as 0.05. In this case, and for the first time using panel data tests, I cannot reject the null hypothesis of a unit root.

The addition of a significant time trend increases the probability of rejection of the null hypothesis. However, for the Prairie provinces, i.e. Manitoba, Saskatchewan and Alberta, addition of a time trend reduces the absolute value of the individual *t*-statistics (see Table 4.3).⁸¹ I investigate the significance of the coefficient on the time trend in the equation for each province: it is only significant (at all levels) for Quebec. When including a time trend, therefore, there is little change in the individual *t*-statistics, but there is a considerable increase in the mean correction factor (a_{NT}), which varies with the inclusion of a time trend. These two effects combine to make the numerator of the standardized *t*-bar test statistic close to zero. I repeat the *t*-bar test, this time using the results from the equation with a time trend for Quebec, and from equation

⁸¹The *t*-statistics for these provinces, in the equation with no time trend, are the largest of the group, and the closest to suggesting rejection of the null hypothesis at the individual level.

(4.22) (excluding a time trend) for all other provincial regressions. The correction factors are adjusted accordingly. In this case, $a_{NT} = -1.54$, $b_{NT} = 0.82$ and the standardized t -bar statistic is 2.45. The data reject the null hypothesis of a unit root.

4.4.2 Industrial data

Panel Dickey-Fuller test

I adjust the manufacturing industry data (as the provincial data) by subtracting the initial value from each time series observation. The resulting equations, for models with and without a time trend, are shown below (standard errors in parentheses),

$$x_{it} = 0.27289 + 0.73456x_{i,t-1} \\ (0.01613) \quad (0.01004)$$

and

$$x_{it} = 0.23388 + 0.78065t + 0.70976x_{i,t-1}. \\ (0.01663) \quad (0.09025) \quad (0.01037)$$

The associated t -statistics for testing $\delta = 1$ are calculated as -26.4 and -28.0 respectively. These clearly exceed the 95% critical value (1.645); the manufacturing industry data also reject the null hypothesis of a unit root.

δ -bar test: base case

As for the provincial data, I subtract the cross-section average from each observation. I regress Δx_{it} and x_{it-1} on 5 lagged first differences. Then I derive $\hat{S}$, estimate equation (4.17) and calculate t_{δ}^* .

The results for both Model 2 (constant, no time trend) and Model 3 (constant plus time trend) are presented in Table 4.4. Once again, the δ -bar test statistics (16.2 for Model 2 and 32.8 for Model 3) exceed the 95% critical value (1.96). The data continue to reject the null hypothesis of a unit root.

δ -bar test: index adjustment

Using the QS [76] dynamic index model, I assume, as for the provincial data, that $M_a = 1$, $M_g = 1$ and $M_b = 2$. Table 4.5 contains the results for Model

	Model 2	Model 3
$\hat{\delta}$	-0.03396090	-0.126373
$\hat{\sigma}_\epsilon$	0.9844993	0.985250
$RSE(\hat{\delta})$	0.006495736	0.011095
t_δ	-5.228184	-11.38994
t_δ^*	16.18465	32.81161

Table 4.4: δ -Bar Test Statistics for Manufacturing Industry Data

	Model 2(1)	Model 2(2)	Model 3(1)	Model 3(2)
$\hat{\delta}$	-0.057845	-0.048936	-0.219748	-0.176447
$\hat{\sigma}_\epsilon$	0.978387	0.978379	0.984247	0.984239
$RSE(\hat{\delta})$	-0.007402	0.006733	0.014215	0.012500
t_δ	-7.814841	-7.267964	-15.45890	-14.11604
t_δ^*	-17.92392	15.74611	46.14207	38.75622

Table 4.5: δ -bar Test Statistics (with Index Adjustment) for Manufacturing Industry Data

2 and Model 3, assuming both one and two common factors (indeces). The critical value is 1.96. Again, the data reject the null hypothesis of a unit root for the one- and two-index cases, in models with and without a time trend.

t -bar test

I subtract the initial cross-section value from all observations and estimate equation (4.22) (model without a time trend). The t -statistics for testing $\delta_i = 0$ are given in Table 4.6. Assuming no time trend, and with $p_i = 12$ for all industries, a_{NT} and b_{NT} are calculated as -1.48 and 0.77 respectively. The standardized t -bar test statistic is 2.17, which again exceeds the 95% critical value of 1.96. The data reject the null hypothesis.

When including a time trend, $p_i = 12$ for all industries, $a_{NT} = -2.13$ and $b_{NT} = 0.61$. The standardized t -bar statistic is given by -0.37 . As for the provincial data, when including a time trend, I can no longer reject the null hypothesis of a unit root. Again, I investigate the significance of the time trend in each individual industry regression: the coefficient is only significantly different from zero for the beverages industry. I compute a mixed test statistic,

Industry	<i>t</i> -Statistic: No Time Trend	<i>t</i> -Statistic: Time Trend
Food	-2.4746	-2.7768
Beverage	-0.54431	-0.79868
Tobacco	-1.1016	-1.7530
Rubber	-2.0967	-3.6837
Leather	-0.77112	-4.0158
Textiles	-0.89989	-2.6638
Clothing	-1.3708	-2.7053
Wood	-1.6962	-1.5991
Furniture	-2.0668	-2.6099
Paper and allied products	-1.0234	-2.2891
Primary metals	-0.65800	-2.5277
Fabricated metals	0.04411	-2.3026
Machinery	-2.9196	-2.9639
Transportation equipment	-0.93501	-2.5770
Electricals and electronics	1.3756	2.4703
Non-metallic mineral products	-2.5187	-2.8119
Petroleum	0.55835	-0.62480
Chemicals	-1.1536	-4.0168
Other	0.35021	-1.4189

Table 4.6: *t*-Statistics (*t*-Bar Test) for Manufacturing Industry Data

	Province	Manufacturing Industry
No time trend	0.51989 (3.78)	0.84999 (3.78)
Index (1) + no time trend	1.6216 (3.78)	2.6638 (3.78)
Index (2) + no time trend	1.7672 (3.78)	2.5636 (3.78)
Time trend	3.4829 (5.34)	1.1083 (5.34)
Index (1) + time trend	1.5653 (5.34)	2.9278 (5.34)
Index (2) + time trend	1.3844 (5.34)	2.8274 (5.34)

Table 4.7: DF Test Statistics for Provincial and Manufacturing Industry Data

using the t -statistics from the equation with a time trend for beverages, and from the equation without a time trend for all other industries. This gives $a_{NT} = -1.52$, $b_{NT} = 0.76$ and a standardized t -bar statistic of 2.27. In this instance, the data reject the null hypothesis.

4.4.3 Dickey-Fuller test

For comparison, I compute a standard Dickey-Fuller test for a unit root, for both data sets. In order to be able to use the test, I average the data over the cross-section. I subtract (i) the simple cross-section average (ii) the average of a single index or (iii) the average sum of two indices, from each cross-section observation, and estimate models with and without a time trend, given by the following equations,

$$(4.25) \quad \Delta y_t = \alpha + \delta y_{t-1} + \sum_{L=1}^p \theta_L \Delta y_{t-L} + \epsilon_t$$

and

$$(4.26) \quad \Delta y_t = \alpha + \delta y_{t-1} + \gamma t + \sum_{L=1}^p \theta_L \Delta y_{t-L} + \epsilon_t.$$

For equation (4.25) (no time trend), I calculate the F -statistic for testing $\alpha = \delta = 0$. For equation (4.26) (including a time trend), I calculate the F -statistic for testing $\delta = \gamma = 0$.

Table 4.7 summarises the results. Critical values are shown in parentheses. I cannot reject the null hypothesis of a unit root for either of the data series. This result holds for both models (with and without a time trend), whether the data are adjusted by subtracting the cross-section average, a single index or a sum of two indices.

4.5 Conclusions

In this Chapter, I present the results from a variety of panel data unit root tests on Canadian data. Each panel unit root test statistic, unlike the Dickey-Fuller test, has a limiting normal distribution. I examine annual data on GPP (1961-1990) and monthly data on manufacturing industry GDP per employee (1974:2 to 1994:11).

To use the standard Dickey-Fuller unit root test, data must be averaged across the cross-section. Using this approach, I cannot reject the null hypothesis of a unit root in either data set. This result holds whether the data are adjusted (for individual effects) using the cross-section average or indeces, and whether or not a time trend is included.

Using a panel Dickey-Fuller test (BM [16]), neither the provincial nor the manufacturing industry data support the hypothesis of a unit root. For the δ -bar test (LL [51]), I adjust the data by using a simple cross-section average, and by allowing for more complex patterns of contemporaneous correlation. In both cases, I again reject the null hypothesis of a unit root. Computation of IPS's [42] t -bar statistic, which explicitly takes account of small-sample bias, gives broadly complementary results. For this test, the results are sensitive to the inclusion of a time trend. However, I show that if, when calculating the test statistic, I only include a time trend in those individual regressions where the coefficient on the time trend is significantly different from zero, then the data once again reject the null hypothesis.

The standard approach to testing for unit roots in panel data, which averages over individual effects, produces misleading results. Furthermore, researchers who first-difference panel data—in the belief that it contains a unit root—needlessly discard useful information.

These results can be related to those of Granger [36], who finds that aggregation of a multivariate time series process can lead to a class of model that has fundamentally different properties. In particular, even if disaggregate series are stationary, their aggregation may contain a unit root. This result explains the apparent inconsistency between the need to first difference the aggregate output data in Chapter 3, and the finding in Chapter 4 that, if cross-sectional information is included, the data appear to be stationary. If, however, the cross-

sectional effects are averaged out, as in the standard Dickey-Fuller approach, the hypothesis that the data contain a unit root cannot be rejected.

Chapter 5

Conclusions

Three essays comprise this thesis. These essays study the relationship between aggregates and disaggregates in the context of Canadian economic growth. My results support the findings in several recent papers that have emphasised the importance of including disaggregate information to explain aggregate activity (see, for example, Durlauf [28], Galor and Zeira [35] and Quah [68]).

In the first essay, Chapter 2, I examine the dynamics of economic growth in Canada, investigating the evolving nature of the underlying cross-section distributions. Previous studies of economic growth and convergence in Canada focus on aggregate measures: a convergence rate of 2% per annum is interpreted as evidence that less well-off regions will, in the long run, catch up to rich regions. However, the methods used in these studies reveal little about fluctuations in the pattern of economic growth within the cross-section. They tell us nothing about the mobility within the cross-section rankings; or about the shape of the cross-section distribution.

Using an econometric framework developed by Quah [69] that incorporates cross-section mobility and shape dynamics, I explore relationships among Canadian disaggregates. I show that the cross-section distribution is unimodal for measures of provincial income per capita (relative to the national average). However, persistence exists in the cross-sectional ordering of relative income levels: poor provinces remain relatively poor; and rich provinces remain relatively rich.

The long-run distribution of gross domestic product per employee across manufacturing sectors displays a bi-modal pattern. Industries tend to remain either at the top or bottom of the cross-section distribution. The long-run bi-modal pattern also appears in Ontario manufacturing industry data, but not for aggregations of eastern or western provinces. Movements in Ontario appear to drive the national pattern, reflecting the relative size of the province and the wide range of manufacturing industries represented there.

For policy makers concerned with provincial income inequalities, this essay provides a mixture of good and bad news. On the one hand, for provincial income per capita, the pattern of economic growth supports the concept of long-run convergence. On the other hand, there is no evidence to suggest that poorer provinces can approximate richer provinces in the short or medium run. Rankings remain unchanged within the cross-section distribution, for provincial and industrial disaggregations.

A comparison of the results for the provincial income per capita data and the manufacturing GDP per employee data suggests that it could be government transfer payments that are behind the unimodal long-run distribution. In the absence of these payments, the evidence suggests that the ergodic distribution is diverging. Hence, as noted by Green [37] in an early study of regional aspects of Canada's economic growth, relying solely on national growth to eliminate the gap between rich and poor provinces is unlikely to be successful. This raises the issue of whether transfer payments can be sustained indefinitely, particularly in the face of increasing constraints on government expenditure. However, this lies beyond the scope of this thesis, but would be an interesting area in which to extend the current work. It appears that it is those factors which determine a region's comparative advantage (such as geographic location, climate, resource endowments, etc.) which determine the steady-state distribution of income across provinces.

Despite directing substantial expenditures towards stimulating particular types of industries in poorer regions, the results suggest that there is little evidence that these industries end up at the top of the cross-section distribution. The pattern of economic growth for industries which receive subsidies does not appear to be distinct from those which do not. Those manufacturing industries at the top of the distribution (refined petroleum and coal products, tobacco, beverages and food) are evenly spread across provinces. The results illustrate the difficulties inherent in targeting assistance towards potential industry "winners".

In the second essay, Chapter 3, I propose a technique to examine the impact of aggregate and disaggregate disturbances on measures of aggregate economic growth and interaction across disaggregates. I use data on provincial income per capita and manufacturing industry GDP per employee. The approach is motivated by studies such as Durlauf [28], that argue that the interaction across dis-

aggregates contains explanatory information for the aggregate, and vice versa. If the disaggregate disturbance can be shown to have some effect on aggregate economic activity, then this strengthens the case for including disaggregate information in an explanation of aggregate growth.

I find that the aggregate disturbance has positive long-run effects on provincial income per capita and GDP per employee, and explains most of their variation. The disaggregate disturbance only matters for fluctuations in aggregate income at business cycle horizons. It has little effect on GDP per employee. However, the disaggregate disturbance has important short and medium-run effects on the interaction measures. In contrast, the aggregate shock contributes little.

I argue in Chapter 2 that an explanation of economic growth and/or convergence should include disaggregate information. The results from Chapter 3 confirm that disaggregate disturbances contain important information for aggregate economic activity at business cycle horizons. Unfortunately, the traditional method of considering convergence concentrates on the explanatory power of aggregate shocks: in the standard neoclassical growth model, long-run changes in output are driven by technology shocks. To gain a more complete picture of convergence the researcher should include disaggregate information. Policy-makers mispecify the expected impact of different macroeconomic policies by ignoring the role played by disaggregate disturbances in the short to medium run. Models which are based on the actions of a representative agent, for example, will miss important information, since they ignore the interaction across disaggregates.

In the third essay, Chapter 4, I present the results from a variety of panel data unit root tests applied to Canadian provincial and manufacturing industry income data. In order to use a standard Dickey-Fuller unit root test, the researcher must average data across the cross-section. Using this approach, I cannot reject the null hypothesis of a unit root for either data set. This result holds whether the data are adjusted for individual effects using the cross-section average or a more complex index of contemporaneous correlation; and whether or not a time trend is included in the estimated equation.

Using a panel Dickey-Fuller test of Breitung and Meyer [16], that removes fixed effects, neither provincial nor manufacturing industry data support the hypothesis of a unit root. For the δ -bar test (Levin and Lin [51]), I adjust

the data by using both a simple cross-section average and a dynamic index representation. In both cases, I again reject the null hypothesis of a unit root. Calculation of Im, Pesaran and Shin's [42] t -bar statistic, which explicitly takes account of small-sample bias, also suggests that in both cases the null hypothesis can be rejected.

The presence of a unit root has important implications for macroeconometric studies. If data contain a unit root, shocks will have permanent effects; where there is no unit root, shocks will have only temporary effects (see, for example, Nelson and Plosser [57]). The results from Chapter 4 suggest that when taking into account cross-section information, there are no permanent shocks in Canadian provincial and manufacturing industry panel income data. Policy makers could misjudge the impact of policy changes if, when estimating the likely effect of a technological disturbance for example, they assume that it will have permanent effects.

The main finding of this thesis is that aggregating data across the cross-section, or ignoring disaggregate information, can produce misleading results. This is illustrated in different ways in the three essays. In Chapter 2, I show that the traditional approach to measuring economic convergence hides the underlying shape and mobility dynamics. In Chapter 3, I demonstrate that disaggregate shocks affect economic growth at business cycle frequencies. However, these shocks are ignored if the cross-sectional dimension is averaged out. Finally, in Chapter 4, I show that aggregating dynamic equations can lead to a model with fundamentally different time-series properties. Policy makers should be alerted to the contribution of disaggregate information and its possible impact on the aggregate. By ignoring such information, they could seriously mispecify their model, and reach inappropriate conclusions.

Bibliography

- [1] Andres, Javier and Ana Lamo (1995): "Dynamics of the Income Distribution Across OECD Countries", CEPR Discussion Paper No. 252, July.
- [2] Banerjee, Anindya, Robin L. Lumsdaine and James H. Stock (1992): "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence", *Journal of Business and Economic Statistics*, Vol. 10, No. 3.
- [3] Barro, Robert J. (1991): "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics*, Vol. 106, No. 2, 407-443.
- [4] Barro, Robert J. and Xavier Sala-i-Martin (1991): "Convergence Across States and Regions", *Brookings Papers on Economic Activity*, Vol. 1, 107-182.
- [5] Barro, Robert J. and Xavier Sala-i-Martin (1992): "Convergence", *Journal of Political Economy*, Vol. 100, No. 2, 223-251.
- [6] Bayoumi, Tamim and Barry Eichengreen (1993): "Shocking Aspects of European Monetary Unification" in F. Torres and F. Giavazzi, eds., *Adjustment and Growth in the European Monetary Union*, Cambridge, UK: Cambridge University Press.
- [7] Bayoumi, Tamim and Barry Eichengreen (1993): "Monetary and Exchange Rate Arrangements for NAFTA", IMF Working Paper, No. 20, March.
- [8] Ben-David, D. (1996): "Trade and Convergence among Countries", *Journal of International Economics*, Vol. 40, 279-298.
- [9] Bernard, Andrew B. and Charles I. Jones (1996): "Productivity Across Industries and Countries: Time Series Theory and Evidence", *The Review of Economics and Statistics*, 135-146.
- [10] Bernard, Andrew B. and Steven N. Durlauf (1996): "Interpreting tests of the convergence hypothesis", *Journal of Econometrics*, Vol. 71, 161-173.
- [11] Berthelemy, J.C. and A. Varoudakis (1996): "Economic Growth, Convergence Clubs, and the Role of Financial Development", *Oxford Economic Papers*, Vol. 48, 300-328.

- [12] Bianchi, Marco (1995): "Testing for Convergence: A Bootstrap Test for Multimodality", Working Paper, Bank of England, May.
- [13] Blanchard, Olivier Jean and Lawrence F. Katz (1992): "Regional Evolutions", *Brookings Papers on Economic Activity*, Vol. 1, 1-76.
- [14] Blanchard, Olivier Jean, and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 79, 655-673.
- [15] Blanchard, O. and D. Quah (1993): "Fundamentalness and the Interpretation of Time Series Evidence: A Reply to Lippi and Reichlin", *American Economic Review*, Vol. 83, 653-658.
- [16] Breitung, J. and W. Meyer (1994): "Testing for unit roots in panel data: are wages on different bargaining levels cointegrated", *Applied Economics*, Vol. 26, 353-361.
- [17] Campbell, J. and P. Perron (1991): "Pitfalls and Opportunities: What macroeconomists should know about unit roots", in *NBER Macroeconomics Annual*, MIT Press.
- [18] Carlino, Gerald and Leonard Mills (1994): "Convergence and the US States: A Time Series Analysis", Federal Reserve Bank of Philadelphia, Working Paper 94-13, July.
- [19] Cass, David (1965): "Optimum Growth in an Aggregative Model of Capital Accumulation", *Review of Economic Studies*, Vol. 32, 233-40.
- [20] Cogley, T. and J.M. Nason (1995): "Output Dynamics in Real Business Cycle Models", *American Economic Review*, Vol. 85, No. 3, 492-511.
- [21] Coulombe, Serge and Frank C. Lee (1994): "Convergence Across Canadian Provinces, 1961 to 1991", *Canadian Journal of Economics*, Vol. XXVIII, No. 4a, 886-898.
- [22] Coulombe, Serge and Frank C. Lee (1995): "Long-run regional growth patterns and economic convergence in Canada", First Draft of paper to be presented at the meetings of the Canadian Economic Association, UQAM, Montreal, 2 June.
- [23] de la Fuente, Angel (1996): "On the Sources of Convergence: A Close Look at the Spanish Regions", Centre for Economic Policy Research, Discussion Paper no. 1543, December.
- [24] den Haan, Wouter J. (1995): "Convergence in stochastic growth models: The importance of understanding why income levels differ", *Journal of Monetary Economics*, Vol. 35, 65-82.
- [25] Dickey, D. and W.A. Fuller (1979): "Distribution of the Estimates for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, Vol. 74, 427-431.

- [26] Dickey, D. and W.A. Fuller (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, Vol. 49, 1057-1072.
- [27] Diebold, F.X. and M. Nerlove (1990): "Unit Roots in Economic Time Series: A Selective Survey", *Advances in Econometrics*, No. 8, 3-69.
- [28] Durlauf, Steven N. (1993): "Nonergodic Economic Growth", *Review of Economic Studies*, Vol. 60, No. 2, 349-366.
- [29] Durlauf, Steven, N. and Paul Johnson (1994): "Nonlinearities in Intergenerational Income Mobility", Working Paper, University of Wisconsin, Economics Department, Madison, WI 53706.
- [30] Enders, W. (1995): *Applied Econometric Time Series*, Wiley Series in Probability and Mathematical Statistics, Wiley: New York.
- [31] Engle, R.F. and J.V. Issler (1995): "Estimating Common Sectoral Cycles", *Journal of Monetary Economics*, Vol. 35, 83-113.
- [32] Evans, P. and G. Karras (1996): "Convergence Revisited", *Journal of Monetary Economics*, Vol. 37, 249-265.
- [33] Evans, P. and G. Karras (1996): "Do Economies Converge? Evidence from a Panel of US States", *The Review of Economics and Statistics*, 384-388.
- [34] Gali, J. (1992): "How Well Does the IS/LM Model Fit Postwar US Data?", *Quarterly Journal of Economics*, Vol. CVII, 709-735.
- [35] Galor, Oded and Joseph Zeira (1993): "Income Distribution and Macroeconomics", *Review of Economic Studies*, Vol. 60, No. 1, 35-52.
- [36] Granger, C.W.J (1980): "Long Memory Relationships and the Aggregation of Dynamic Models", *Journal of Econometrics*, Vol. 14, 227-238.
- [37] Green, Alan G. (1971): *Regional Aspects of Canada's Economic Growth*, University of Toronto Press, Toronto and Netherlands.
- [38] Hamilton, James D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle" *Econometrica*, Vol. 57, No. 2, 357-384.
- [39] Helliwell, John F. (1992): "Trade and Technical Progress", National Bureau of Economic Research Working Paper No. 4226.
- [40] Helliwell, John F. (1994): "Convergence and Migration among Provinces", Working Paper No. 94-05, Department of Economics, Dalhousie University, July.
- [41] Helliwell, John F. and Alan Chung (1992): "Convergence and Growth Linkages between North and South", National Bureau of Economic Research Working Paper No. 3948.

- [42] Im, K.S., M.H. Pesaran and Y. Shin (1995): "Testing for Unit Roots in Heterogeneous Panels", DAE Working Paper No. 9526, University of Cambridge.
- [43] Islam, Nazrul (1995): "Growth Empirics: A Panel Data Approach", *The Quarterly Journal of Economics*, November, 1127-1170.
- [44] Koopmans, T.C. (1965): "On the Concept of Optimal Economic Growth", in *The Economic Approach to Development Planning*, Pontifical Academy of Sciences, Amsterdam: North-Holland.
- [45] Krugman, Paul (1993): "Lessons of Massachusetts for EMU", in F. Torres and F. Giavazzi, eds., *Adjustment and Growth in the European Monetary Union*, Cambridge, UK: Cambridge University Press.
- [46] Lee, Frank C. (1996): "Convergence in Canada?", *Canadian Journal of Economics*, Vol. XXIX, Special Issue, April, S331-S336.
- [47] Lee, Frank C. and Serge Coulombe (1994): "Regional Productivity Convergence in Canada", mimeo, Department of Finance, Federal Government of Canada, September.
- [48] Lee, Kevin, M. Hashem Pesaran and Ron Smith (1995): "Growth and Convergence: A Multi-Country Empirical Analysis of the Solow Growth Model", Department of Applied Economics Working Paper No. 9531, University of Cambridge.
- [49] Lefebvre, Mario. (1994): "Les Provinces Canadiennes et al Convergence: Une Evaluation Empirique", Bank of Canada Working Paper No. 94-10, November.
- [50] Lefebvre, Mario and Stephen S. Poloz (1996): "The Commodity-Price Cycle and Regional Economic Performance in Canada", Bank of Canada Working Paper No. 96-12, September.
- [51] Levin, Andrew, and Chien-Fu Lin (1993): "Unit Root Tests in Panel Data: New Results", Discussion Paper No. 93-56, University of California, San Diego, December.
- [52] Lippi, M. and L. Reichlin (1993): "A Note on Measuring the Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, Vol. 83, 644-52.
- [53] Long, John B. and Charles I. Plosser (1983): "Real Business Cycles", *Journal of Political Economy*, Vol. 91, No. 1, 39-69.
- [54] MacDonald, R. (1996): "Panel Unit Root Tests and Real Exchange Rates", *Economics Letters*, Vol. 50, 7-11.
- [55] Mankiw, N.G., D. Romer and D. Weil (1992): "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, Vol. 107, 407-37.

- [56] Melvin, James R. (1987): "Regional Inequalities in Canada: Underlying Causes and Policy Implications", *Canadian Public Policy*, Vol. XIII, No. 3, 304-317.
- [57] Nelson, C.R. and C. I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, Vol. 10, 139-162.
- [58] Otto, Glenn and Tony Wirjanto (1990): "Seasonal Unit-Root Tests on Canadian Macroeconomic Time Series", *Economics Letters*, Vol. 34, 117-120.
- [59] Pagan, Adrian (1984): "Econometric Issues in the Analysis of Regressions with Generalized Regressors", *International Economic Review*, Vol. 25, No. 1, 221-248.
- [60] Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, Vol. 57, 1361-1401.
- [61] Pesaran, M. H. and R. Smith (1995): "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", forthcoming in *Journal of Econometrics*.
- [62] Pesaran, M. H., R. Smith and K. Im (1995): "Dynamic Linear Models for Heterogeneous Panels", DAE Working Paper No. 9503, University of Cambridge, forthcoming in Matyas, L. and P. Sevestre, eds. *Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd Edition, Kluwer Academic Publishers.
- [63] Phillips, P. and P. Perron (1988): "Testing for a Unit Root in Time Series Regression", *Biometrika*, Vol. 75, 335-346.
- [64] Quah, Danny (1992a): "Empirical Cross-Section Dynamics in Economic Growth", LSE Working Paper, October.
- [65] Quah, Danny (1992b): "International Patterns of Growth: I. Persistence in Cross-Country Disparities", LSE Working Paper, November.
- [66] Quah, Danny (1992c): "International Patterns of Growth: II. Persistence, Path Dependence, and Sustained Take-Off in Growth Transition", LSE Working Paper, November.
- [67] Quah, Danny (1993): "Galton's Fallacy and Tests of the Convergence Hypothesis", *The Scandinavian Journal of Economics*, Vol. 95, No. 4, 427-444.
- [68] Quah, Danny (1994): "One Business Cycle and One Trend From (Many,) Many Disaggregates", *European Economic Review*, Vol. 38, 605-613.
- [69] Quah, Danny (1994a): "One Business Cycle and One Trend From (Many,) Many Disaggregates", *European Economic Review*, Vol. 38, 605-613.

- [70] Quah, Danny (1994c): "Ideas Determining Convergence Clubs", mimeo, LSE Economics Department, FMG, and CEPR, September.
- [71] Quah, Danny (1994d): "Exploiting Cross Section Variation for Unit Root Inference in Dynamic Data", *Economics Letters*, Vol.44, 9-19.
- [72] Quah, Danny T. (1995): "Aggregate and Regional Disaggregate Fluctuations", mimeo, LSE Economics Department, March.
- [73] Quah, Danny T. (1996): "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics", *The Economic Journal*, Vol. 106, No. 437, 1045-1055.
- [74] Quah, Danny T. (1996a): "Twin Peaks: Growth and Convergence in Models of Distribution Dynamics", *The Economic Journal*, Vol. 106, No. 437, 1045-1055.
- [75] Quah, Danny (1996b): "Empirics for Economic Growth and Convergence", *European Economic Review*, forthcoming.
- [76] Quah, Danny and Thomas J. Sargent (1993): "A Dynamic Index Model for Large Cross Sections", in *Business Cycles, Indicators, and Forecasting*, Eds. J.H. Stock and M.W. Watson, The University of Chicago Press, Chicago, 1993.
- [77] Sala-i-Martin, Xavier (1994): "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence", CEPR Discussion Paper No. 1075, November.
- [78] Shapiro, M. and M. Watson (1988): "Sources of Business Cycle Fluctuations", *NBER Macroeconomic Annual*, Cambridge, Mass.: MIT Press, 3, 111-148.
- [79] Silverman, B.W. (1986): *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability 26, Chapman and Hall, London.
- [80] Solow, Robert M. (1956): "A Contribution to the Theory of Economic Growth", *The Quarterly Journal of Economics*, Vol. LXX, 65-94.
- [81] Watson, M.W. and R.F. Engle (1983): "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models", *Journal of Econometrics*, Vol. 23, No. 23, 385-400.
- [82] White, K.J. (1978): "A General Computer Program for Econometric Methods - SHAZAM", *Econometrica*, 239-240.
- [83] Zivot, Eric and Donald W.K. Andrews (1992): "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis", *Journal of Business & Economic Statistics*, Vol. 10, No.3, 251-270.

Appendices

A The Data

A.1 Data Sources

Variable	Periods	Source (CANSIM Matrix No.)
Provincial Data		
Population	1926-1993	6967, 6968-6977
Total employment	1966-1993	6967, 6968-6977
Labour force	1966-1993	6967, 6968-6977
GDP in constant dollars	1984-1993	7904-7913
GDP in current dollars	1984-1991	7360-7369
GDP	1961-1993	6967, 6968-6977
GDP/person	1961-1993	6967, 6968-6977
Wages, salaries & SLI (GDP-based)	1961-1993	2611-2619, 6949
Wages, salaries & SLI (Sources & disp. pers. inc)	1961-1993	5080-5097, 6965
Average hourly earnings (1986=100)	1981-1993	6967, 6968-6977
Personal income	1926-1993	6967, 6968-6977
Personal income/person	1926-1993	6967, 6968-6977
Disposable income	1926-1993	6967, 6968-6977
Disposable income/person	1926-1993	6967, 6968-6977
CPI	1978-1993	6967, 6968-6977
Implicit price index	1971-1993	6967, 6968-6977

Variable	Periods	Source (CANSIM Matrix No.)
Industry Data		
All employees, all sizes	1983:1-1995:1	4285
Employment index	1961:1-1983:3	1432
Employees	1961-1992	7951-7964
	1970-1982	7505-7689
	1981-1993	5380-5586, 6849-6899
Total salaries & wages	1981-1993	5380-5586, 6849-6899
Wages, salaries & SLI (15 sectors)	1926-1992	6655
Average hourly earnings	1961:1-1983:3	1435
	1983:1-1995:1	4296, 4298
GDP at factor cost (1986 prices)	1961-1994	4670
GDP at factor cost (1986 prices)	1961:1-1995:1	4674
Real GDP/person at work	1961-1991	7920
Real GDP/person hour	1961-1991	7921
Index real GDP/person	1961-1991	7926
Index real GDP/person/hr	1961-1991	7927

Variable	Periods	Source (CANSIM Matrix No.)
Provincial Industry Data		
Total employees (21 manufacturing ind.)	1970-1982 1981-1993	7482-7501 5379, 5401, 5406, 5409, 5413, 5419, 5424, 5429, 5439, 5458, 5473, 5482, 5496, 5504, 5515, 5540, 5548, 5567, 6848, 6865, 6869, 6883
All employees, all sizes	1983:1-1995:1	4299, 4313, 4327, 4341, 4355, 4369, 4383, 4397, 4411, 4425
Average hourly earnings	1961:1-1983:3 1983:1-1995:1	1445, 1455, 1460, 1465, 1470, 1480, 1485, 1490, 1495 4310, 4324, 4338, 4352, 4366, 4380, 4394, 4408, 4422, 4436
Salaries and wages (21 manufacturing ind.)	1970-1982 1981-1993	7482-7501 5379, 5401, 5406, 5409, 5413, 5419, 5424, 5429, 5439, 5458, 5473, 5482, 5496, 5504, 5515, 5540, 5548, 5567, 6848, 6865, 6869, 6883
GDP at factor cost (current prices)	1971-1984 1984-1993	7871-7880 7360-7369

A.2 Manufacturing Industries

Food
Beverages
Tobacco products
Rubber products
Plastic products
Leather and allied products
Primary textile and textile products
Clothing
Wood
Furniture and fixtures
Paper and allied products
Printing, publishing and allied
Primary metal
Fabricated metal products
Machinery
Transportation equipment
Electrical and electronic products
Non-metallic mineral products
Refined petroleum and coal
Chemical and chemical products
Other manufacturing

B Fractile Transition Probability Matrices

Table B1: WSSLI per Employee by Province
1-year transitions

(Number)	Quantile				
	0.20	0.40	0.60	0.80	1.00
(52)	1.00				
(52)		0.75	0.25		
(52)		0.25	0.73	0.02	
(52)			0.02	0.90	0.08
(52)				0.08	0.92
1-year transitions					
(Number)	Quantile				
	0.25	0.50	0.75	1.00	
(52)	1.00				
(78)		0.86	0.14		
(52)		0.21	0.71	0.08	
(78)			0.05	0.95	

Table B2: GDP per Employee, 21 Manufacturing Industries
1-month transitions

(Number)	Quantile				
	0.20	0.40	0.60	0.80	1.00
(1620)	0.95	0.05			
(1620)	0.07	0.86	0.09		
(1620)		0.09	0.81	0.09	
(1620)			0.10	0.82	0.08
(1762)			0.01	0.08	0.92
1-month transitions					
(Number)	Quantile				
	0.25	0.50	0.75	1.00	
(2025)	0.95	0.05			
(2025)	0.05	0.86	0.09		
(2025)		0.09	0.82	0.08	
(2167)			0.08	0.92	

C Further Estimations

In this Appendix, I present the results from estimation of transition probability matrices for data on provincial income per capita growth, on wages, salaries and supplementary labour income (WSSLI) per capita, and on employment per capita. I also examine GDP per person at work and GDP per person hour for seven different industry sectors across Canada.

C.1 Provincial Income Growth per Capita

First, I estimate transition probability matrices for provincial personal income *growth* per capita. Figure C1 shows the annual growth in provincial personal income per capita, relative to the national average, for 1927-1993. Saskatchewan has the most volatile growth rate; by comparison, most other provinces vary relatively little.⁸²

Table C1: Transition Probability Matrix—Income Growth by Province
1-year transitions

(Number)	Upper endpoint				
	-0.022	-0.004	0.008	0.026	0.699
(128)	0.23	0.11	0.07	0.19	0.41
(126)	0.11	0.29	0.29	0.17	0.13
(123)	0.10	0.24	0.28	0.28	0.11
(125)	0.14	0.21	0.26	0.24	0.15
(126)	0.42	0.17	0.12	0.11	0.17
Steady-state distribution	0.198	0.203	0.204	0.199	0.196
5-year transitions					
(Number)	Upper endpoint				
	-0.022	-0.004	0.008	0.026	0.699
(125)	0.25	0.13	0.18	0.14	0.30
(113)	0.19	0.22	0.28	0.20	0.11
(107)	0.10	0.28	0.27	0.21	0.14
(112)	0.10	0.26	0.24	0.31	0.09
(121)	0.31	0.17	0.10	0.17	0.26
Steady-state distribution	0.181	0.216	0.222	0.209	0.172

⁸²The variance for Saskatchewan is more than six times the provincial average. Ontario and Quebec have the least variable income per capita growth over the sample period.

Transition probability matrices for these data (Table C1) show much more mobility throughout the cross-section distribution than was evident in the levels data—there is no discernible pattern of persistence. Each panel in Table C1 illustrates that it is quite common for economies to reach any other state from their initial position. The minimum entries in the off-diagonals are 7% and 10% respectively in each panel; in some cases, the largest entries are off the diagonal.⁸³ In other words, provinces do not consistently exhibit high or low growth rates (relative to the national average) in income per capita. The ergodic distribution displays convergence towards the middle of the distribution; it is unimodal and symmetric.

Plotting the quantiles (Figure C2) illustrates that much of the volatility occurs in the first half of the sample period, and is contained in the end quantiles. Most of the variance in the top quantile is attributable to fluctuations in income growth in Saskatchewan and Manitoba. In both provinces, a relatively large proportion of provincial GDP is attributable to agriculture and related products.⁸⁴

C.2 Provincial WSSLI per Capita

Annual data on wages, salaries and supplementary labour income (WSSLI) per capita are available for the period 1961-1991. Figure C3 reveals that, relative to the national average, WSSLI per capita has remained fairly steady over the sample period.

Table C2 shows transition probability matrices for these data. The degree of mobility is very similar to that found in the personal income per capita data. Note, however, the stronger persistence in states 2 and 3. This links with the patterns displayed by the ergodic distributions—both reach a peak in state 3.⁸⁵ There appears to be a general tendency for WSSLI per capita to move towards the middle of the cross-section distribution and stay there. Figure C4, which graphs the quantiles, shows some reduction in dispersion (or the “shape”) of WSSLI per capita, relative to the national average. In particular, the bottom quantile is increasing faster than the other quantiles, while, at the same time, the top quantile is falling over the sample period.

⁸³ Estimation with a ten-year transition period did not change the results significantly.

⁸⁴ In Ontario, less than 10% of GDP is generated by agriculture and related products; in Manitoba and Saskatchewan, this figure rises to more than 25%.

⁸⁵ A six-year transition period gives a similar pattern.

Table C2: Transition Probability Matrix—WSSLI per Capita, by Province
1-year transitions

(Number)	Upper endpoint				
	-0.283	-0.116	0.082	0.198	0.425
(62)	0.95	0.05			
(59)		0.92	0.08		
(58)		0.07	0.88	0.05	
(60)			0.08	0.90	0.02
(61)				0.03	0.97
Steady-state distribution	0.000	0.296	0.364	0.226	0.115

(Number)	Upper endpoint				
	-0.291	-0.117	0.085	0.201	0.425
(60)	0.85	0.15			
(48)	0.04	0.83	0.13		
(52)		0.12	0.77	0.12	
(53)			0.19	0.75	0.06
(57)				0.11	0.89
Steady-state distribution	0.082	0.296	0.321	0.196	0.105

C.3 Provincial Employment per Capita

Annual employment per capita, relative to the national average, for 1966-1992, is plotted in Figure C5. Cross-section dispersion has changed little over the sample period: the range is largely constant. There is also limited movement within the cross-section distribution: most provinces maintain a particular rank within the cross-section (Alberta and Ontario at the top, Newfoundland and PEI at the bottom).

This is confirmed by results from estimation of transition probability matrices (Table C3). The probability of remaining in the bottom (top) quantile is at least 91% (96%) with a three-year transition period.⁸⁶ The ergodic distributions, where they exist, suggest convergence towards the bottom half of the distribution.

⁸⁶Even with a six-year transition period, these probabilities remain at 90% and 95% respectively.

Table C3: Transition Probability Matrix—Employment per Capita, Relative to the National Average, by Province

1-year transitions					
(Number)	Upper endpoint				
	-0.098	-0.017	0.049	0.097	0.194
(51)	0.92	0.08			
(49)	0.06	0.94			
(50)			0.90	0.10	
(49)			0.10	0.86	0.04
(51)				0.06	0.94
Steady-state distribution	0.438	0.562	0.000	0.000	0.000
3-year transitions					
(Number)	Upper endpoint				
	-0.100	-0.022	0.049	0.099	0.194
(46)	0.91	0.09			
(42)	0.02	0.98			
(41)			0.85	0.15	
(46)			0.20	0.80	
(45)				0.04	0.96
Steady-state distribution	n/a	n/a	n/a	n/a	n/a

C.4 Industrial GDP per Employee

Annual index data on GDP per person at work and GDP per hour worked are available for seven different industry sectors for 1961-1994.⁸⁷ Figures C6 and C7 show that the cross-section distributions of each measure are similar: values for agriculture (No. 1) are highly variable; those for communications (No. 5) show a steady rise; but productivity appears to have fallen in the service sector (No. 7).

Transition probability matrices (Tables C4 and C5) show similar degrees of mobility within the cross-section. Most persistence is at the ends of the distribution, particularly the top end; mobility is highest in the middle. Mobility is higher than for provincial disaggregation using both WSSL and employment per capita data. Ergodic distributions are unimodal for both measures.

⁸⁷ The sectors studied are agriculture, manufacturing, construction, transportation and storage, communications, wholesale and retail trade, and services.

Table C4: Transition Probability Matrix—GDP per Person at Work,
Relative to the Industry Average, by Industry Sector
1-year transitions

(Number)	Upper endpoint				
	-0.104	-0.027	0.011	0.097	0.499
(44)	0.77	0.16	0.07		
(46)	0.15	0.39	0.30	0.15	
(44)	0.02	0.32	0.34	0.30	0.02
(45)		0.13	0.31	0.49	0.07
(45)		0.02		0.09	0.89
Steady-state distribution	0.165	0.214	0.221	0.222	0.178

3-year transitions					
(Number)	Upper endpoint				
	-0.101	-0.027	0.013	0.101	0.499
(41)	0.61	0.20	0.10	0.10	
(39)	0.13	0.33	0.23	0.28	0.03
(42)	0.05	0.29	0.33	0.26	0.07
(41)	0.07	0.27	0.37	0.24	0.05
(40)			0.05	0.20	0.75
Steady-state distribution	0.151	0.240	0.241	0.229	0.138

Table C5: Transition Probability Matrix—GDP per Person Hour at Work,
Relative to the Industry Average, by Industry Sector
1-year transitions

(Number)	Upper endpoint				
	-0.088	-0.028	0.011	0.009	0.422
(44)	0.68	0.25	0.05	0.02	
(46)	0.22	0.39	0.28	0.11	
(45)	0.04	0.27	0.38	0.31	
(44)	0.02	0.09	0.32	0.50	0.07
(45)				0.09	0.91
Steady-state distribution	0.187	0.206	0.219	0.220	0.168

3-year transitions					
(Number)	Upper endpoint				
	-0.087	-0.028	0.009	0.106	0.422
(40)	0.57	0.20	0.15	0.08	
(40)	0.13	0.40	0.20	0.25	0.03
(41)	0.12	0.29	0.27	0.27	0.05
(42)	0.10	0.12	0.40	0.33	0.05
(40)			0.08	0.17	0.75
Steady-state distribution	0.188	0.226	0.240	0.232	0.114

Figure C1
Growth of Provincial Personal Income per Capita (Relative to National Average)
1927-1993

Figure C2
Quantiles for Growth of Provincial Personal Income per Capita (1929-1992)

Figure C3
Log of Provincial WSSLI per Capita (Relative to National Average) 1961-1992

Figure C4
Quantiles for Log of Provincial WSSL per Capita (1963-1991)

Figure C5
Provincial Employment per Capita (Relative to National Average) 1966-1993

Figure C6
Index of GDP per Employee 1986=100 (Relative to National Average)
1961-1994

Figure C7
Index of GDP per Person Hour Worked 1986=100 (Relative to National Average)
1961-1994

D Tests For Robustness

Table D1: Transition Probability Matrix—Income by Province, Relative to Ontario

1-year transitions					
(Number)	Upper endpoint				
	-0.543	-0.376	-0.258	-0.134	0.126
(118)	0.94	0.05		0.01	
(118)	0.03	0.82	0.11	0.04	
(115)		0.09	0.80	0.10	0.02
(116)	0.01	0.03	0.09	0.77	0.09
(114)			0.01	0.10	0.89
Steady-state distribution	0.114	0.186	0.223	0.232	0.245
3-year transitions					
(Number)	Upper endpoint				
	-0.545	-0.381	-0.258	-0.130	0.126
(117)	0.86	0.13		0.01	
(117)	0.03	0.71	0.21	0.04	0.02
(107)	0.01	0.13	0.66	0.18	0.02
(108)	0.02		0.14	0.69	0.16
(105)		0.02	0.03	0.13	0.82
Steady-state distribution	0.081	0.156	0.228	0.266	0.270

Table D2: Transition Probability Matrix—Income by Province (subtracting Ontario)

1-year transitions					
(Number)	Upper endpoint				
	-1.561	-1.203	-0.811	-0.519	0.891
(117)	0.85	0.14	0.01		
(114)	0.12	0.78	0.08	0.02	
(117)		0.10	0.72	0.15	0.03
(117)		0.01	0.17	0.74	0.08
(116)			0.03	0.09	0.89
Steady-state distribution	0.184	0.218	0.204	0.197	0.197
3-year transitions					
(Number)	Upper endpoint				
	-1.561	-1.192	-0.808	-0.513	0.891
(116)	0.75	0.23	0.02		
(105)	0.17	0.66	0.16		0.01
(112)	0.01	0.14	0.58	0.22	0.04
(114)	0.01	0.01	0.21	0.66	0.11
(107)		0.01	0.06	0.11	0.82
Steady-state distribution	0.160	0.211	0.220	0.209	0.201

Table D3: Transition Probability Matrix—Income by Province, Aggregating some Provinces

1-year transitions					
(Number)	Upper endpoint				
	-0.102	0.045	0.137	0.260	0.453
(80)	0.91	0.08	0.01		
(78)	0.06	0.67	0.24	0.03	
(77)	0.01	0.26	0.68	0.05	
(79)		0.01	0.09	0.84	0.06
(76)				0.09	0.91
Steady-state distribution	0.212	0.241	0.239	0.183	0.126
3-year transitions					
(Number)	Upper endpoint				
	-0.112	0.043	0.137	0.255	0.453
(76)	0.92	0.07	0.01		
(74)	0.03	0.61	0.32	0.04	
(69)	0.03	0.30	0.54	0.13	
(75)		0.04	0.17	0.68	0.11
(78)		0.01		0.15	0.83
Steady-state distribution	0.178	0.250	0.253	0.194	0.124

Table D4: Transition Probability Matrix—GDP per Employee,
Manufacturing Industries, Ontario

1-year transitions

(Number)	Upper endpoint				
	-0.355	-0.120	0.015	0.204	0.970
(77)	0.81	0.16		0.01	0.03
(79)	0.18	0.62	0.18	0.01	0.01
(77)		0.18	0.55	0.25	0.03
(80)	0.01	0.01	0.23	0.63	0.13
(86)	0.02		0.03	0.08	0.86
Steady-state distribution	0.202	0.176	0.181	0.187	0.255

3-year transitions

(Number)	Upper endpoint				
	-0.350	-0.119	0.026	0.211	0.802
(66)	0.70	0.18	0.03	0.02	0.08
(66)	0.23	0.48	0.18	0.06	0.05
(71)	0.03	0.25	0.45	0.24	0.03
(67)	0.03	0.06	0.21	0.57	0.13
(66)	0.02		0.09	0.11	0.79
Steady-state distribution	0.184	0.179	0.186	0.197	0.253

Table D5: Transition Probability Matrix—GDP per Employee,
Manufacturing Industries, Western Provinces

1-year transitions

(Number)	Upper endpoint				
	-0.601	-0.341	-0.117	0.301	2.000
(60)	0.77	0.17	0.02	0.02	0.03
(72)	0.24	0.57	0.17	0.01	0.01
(70)	0.04	0.21	0.60	0.13	0.01
(75)		0.03	0.16	0.63	0.19
(75)	0.03	0.01	0.01	0.20	0.75
Steady-state distribution	0.259	0.203	0.173	0.179	0.187

3-year transitions

(Number)	Upper endpoint				
	-0.572	-0.308	-0.082	0.328	1.616
(60)	0.78	0.07	0.05	0.05	0.05
(51)	0.16	0.51	0.33		
(62)		0.37	0.50	0.08	0.05
(64)	0.02	0.03	0.11	0.56	0.28
(58)	0.05	0.02	0.02	0.31	0.60
Steady-state distribution	0.206	0.202	0.205	0.196	0.190

Table D6: Transition Probability Matrix—GDP per Employee,
Manufacturing Industries, Eastern Provinces
1-year transitions

(Number)	Upper endpoint				
	-0.504	-0.203	-0.008	0.208	1.476
(70)	0.87	0.13			
(69)	0.12	0.62	0.23	0.01	0.01
(71)		0.23	0.49	0.23	0.06
(71)			0.23	0.59	0.18
(89)		0.01	0.07	0.12	0.80
Steady-state distribution	0.169	0.187	0.204	0.194	0.246

3-year transitions					
(Number)	Upper endpoint				
	-0.504	-0.198	-0.001	0.217	1.476
(62)	0.77	0.19		0.03	
(61)	0.11	0.49	0.26	0.08	0.05
(61)		0.21	0.49	0.21	0.08
(62)		0.03	0.23	0.45	0.29
(64)		0.03	0.03	0.19	0.75
Steady-state distribution	0.075	0.148	0.198	0.225	0.355

Figure D1
Quantiles for Log of Provincial Personal Income per Capita,
Divided by Ontario Data (1928-1993)

Figure D2
Quantiles for Log of Provincial Personal Income per Capita,
Subtracting Ontario Data (1928-1993)

Figure D3
Quantiles for Log of Aggregated Provincial Personal Income per Capita
(1928-1991)

Figure D4
Quantiles for Log of Manufacturing Industry GDP per Employee:
Ontario (1973-1990)

Figure D5
Quantiles for Log of Manufacturing Industry GDP per Employee:
Western Provinces (1973-1990)

Figure D6
Quantiles for Log of Manufacturing Industry GDP per Employee:
Eastern Provinces (1973-1990)

E Results Using Coulombe and Lee Data

This Appendix summarizes the results arising from studying the mobility dynamics of data used by Coulombe and Lee in their various studies of convergence across Canadian provinces (see [21], [22] and [47]).

E.1 Data

In their first paper, Coulombe and Lee [21] draw a distinction between provincial product and income data for the period 1961-1990. They use the measures shown in Table E1. In their second paper, [47], Coulombe and Lee examine convergence in productivity, using the variables listed in Table E2 below for the period 1966-1990. The most recent paper, [22], uses four measures of per capita income over the longer period 1926-1991 (see Table E3).

From these papers, Coulombe and Lee conclude that there is evidence of convergence in both product and income measures, with the speed of convergence increasing as disaggregation moves from a per capita, to a per worker, to a per hour basis. Most of the convergence occurred during the period 1950-1977. Regional shocks in the 1930s and 1940s outweighed any evidence for β -convergence prior to 1950.

Regional disparities in unemployment rates are the most significant factors in slowing down regional labour market adjustments, which, in turn, hinder convergence in living standards. Provincial price deflators diminish the speed

Table E1: Data Used in Coulombe and Lee [21]

Variable	Definition
GPPRO	Gross provincial product at factor cost per capita, provincial price indices
GPPNAT	Gross provincial product at factor cost per capita, national price index
EI	Earned income per capita
PIT	Personal income minus government transfers per capita
PI	Personal income per capita
PDI	Personal disposable income per capita

Table E2: Data Used in Coulombe and Lee [47]

Variable	Definition
GPPC	Gross provincial product at factor cost per capita
GPPW	Gross provincial product at factor cost per worker
GPPH	Gross provincial product at factor cost per hour worked
EIC	Earned income per capita
EIW	Earned income per worker
EIH	Earned income per hour worked

Table E3: Data Used in Coulombe and Lee [22]

Variable	Definition
PI	Personal income per capita
PIT	Personal income minus government transfer payments per capita
PDI	Personal disposable income per capita
EI	Earned income per capita

of convergence of regional output but do not appear to affect the rate of convergence of income measures. This difference is attributed to the uneven distribution of economic activity across Canada. Consumption patterns appear to be similar.

E.2 Mobility Dynamics

I estimate transition probability matrices for each of the variables listed above, and present the results for the different variables in turn.

E.2.1 Output and income per capita

The first variable I study is gross provincial product (GPP) per capita, deflated by (i) a national (GPPNAT) and (ii) provincial price indices (GPPRO) (see Coulombe and Lee [21]).

Alberta has the highest average GPP per capita (relative to the national average) over the sample period, followed by Ontario, British Columbia and Saskatchewan. The Maritime provinces are at the bottom of the ranking.

Provincial price deflators increase cross-section mobility and tend to smooth the output data.⁸⁸ There is a slight narrowing of the range between the ends of the cross-section distribution over the sample period. This range is smaller for the nationally deflated data. This supports the Coulombe and Lee [21] finding that provincial price deflators reduce the speed of convergence.

Estimation of transition probability matrices for each output measure suggest convergence, both in terms of a reduction in cross-section dispersion (shape)

⁸⁸This suggests noise in the provincial price data.

and within the cross-section distribution (mobility). It appears that convergence is towards state 2.

Most of the income measures exhibit high mobility in state 2 and high persistence at the top of the cross-section distribution. The top is occupied mainly by Ontario; the volatile agricultural economy of Saskatchewan influences mobility further down the distribution. Most quantiles dip in 1980, except the fifth quantile—which includes Alberta data—which increases in the same period. This reflects the recession that hit Ontario particularly hard and reveals an interesting distinction between the output and income data. Alberta had the highest output per capita (relative to the national average) over the sample period, but for the income data, it drops to third place, behind Ontario and British Columbia. The boost to provincial output from the oil boom was not immediately transmitted into higher income.

Results for per capita earned income (EI) and per capita personal income minus government transfers (PIT) are very similar. Provincial rankings are identical. The quantiles exhibit a reduction in dispersion over time. Therefore, the addition of interest and dividend payments does not appear to affect the pattern of convergence—a result consistent with Coulombe and Lee [21]. The ergodic distributions peak in state 4.

The addition of government transfer payments (PI) pushes the ergodic peak up the cross-section distribution. In comparison with PIT, values for the top two quantiles are lower, and the bottom three quantiles higher: transfer payments help to reduce the disparity in personal incomes. The Maritime provinces, led by Newfoundland, received the highest value per capita of government transfer payments to persons over the sample period. PI shows more mobility at the bottom end of the cross-section distribution than the other income measures. However, contrary to expectations, this increased mobility is down the distribution rather than up it. Personal disposable income per capita (PDI) data—which removes direct taxes—covers a narrower range, but exhibits similar patterns of mobility.⁸⁹

E.2.2 Productivity measures

I now examine the different measures of provincial productivity—output and income per capita, per worker, and per hour worked—used by Coulombe and Lee in [47].

Comparing first the output measures, gross provincial product per worker (GPPW) displays the most mobility, particularly at the top end of the cross-section distribution. Both GPP per capita (GPPC) and GPPW converge towards the middle of the cross-section distribution in the long term.

Ranking the provinces by the average value of GPP per hour (GPPH) (relative to the national average) over the sample period, places the two western-most provinces at the top, Newfoundland in the middle, and the remaining Maritime provinces at the bottom. This suggests that, although there is high unemployment in Newfoundland, those industries that do employ workers are highly

⁸⁹Newfoundland moves up to 8th position—from 9th—using this measure.

“productive”, in terms of output per hour worked.⁹⁰ GPPH exhibits more mobility at the bottom of the cross-section distribution, and is the only variable to display divergence across the cross-section. This divergence could be caused by the size of the distance between quantiles. The end quantiles are further away from the others. Alternatively, it could be attributed to the type and range of industries represented in the different provinces (see Chapter 2 for a fuller discussion).

Rankings of provinces vary across the different income variables: Ontario, Alberta and British Columbia share the top for all three measures, but Saskatchewan appears near the bottom for earned income per worker (EIW) and earned income per hour (EIH), whilst Newfoundland moves up the distribution to 7th place for EIW and to 4th place using EIH data. This reflects the impact of unemployment and the type of industries in each province on the results. The Saskatchewan result can be attributed to the agricultural base of the province. This economy is heavily dependent upon agricultural prices and the success of harvests. As a consequence, incomes tend to be highly variable and largely seasonal. This volatility is reflected in the transition probability matrices. All the income data exhibit high mobility in state 2. Probabilities and persistence are high at the top end of the cross-section distribution.

Dispersion has declined in the earned income per capita (EIC) and EIH data, but less so in the EIW series. Further investigation is required to determine whether this reduction arises from a tendency for different industries to locate more widely (rather than concentrating production in particular provinces), or if it reflects a convergence in productivity across industries. The EIW results could be attributed (as in Coulombe and Lee [21]), to consistently disparate unemployment rates across provinces.

Ergodic distributions for the EIC and EIW series are similar, peaking in states 3 and 4 respectively.

E.2.3 Longer sample period

Personal income per capita (PI) and personal disposable income per capita (PDI) produce very similar transition probability matrices and ergodic distributions. In the long term, observations within the cross-section distribution converge to the middle of the distribution.

E.3 Conclusions

Mobility dynamics, for both output and income variables—per capita and per worker—suggest convergence both across (shape) and within (mobility) the cross-section distributions. There is high persistence at the top of the distribution, particularly for the income measures of economic growth. In particular, Ontario remains consistently at the peak of the distribution. High mobility

⁹⁰In 1994, Newfoundland's unemployment rate was 20.4%, while the Canadian average was 10.4%. The food and printing and publishing industries are the most important manufacturing industries in Newfoundland (in terms of contribution to GDP). GDP per employee in food manufacturing exceeds the average across all manufacturing industries.

is evident lower down the cross-section distribution, and is attributable to the uncertain, volatile incomes those provinces generate.

The different output measures of productivity used in [47] suggest different long run patterns of convergence. Both GPPC and GPPW exhibit convergence, but GPPH indicates divergence. The income data all suggest convergence in the long term.

These results indicate that there does not appear to be a direct one-to-one relationship between output and income measures. Measures of income have more of an opportunity to converge. Government transfers to persons and migration tend to be the dominant sources that reduce dispersion. Measures of output are more heterogeneous, and possess relatively wide variations across provinces. This reflects the diverse natural resources, climates and industries of the different provinces.

These findings mainly support Coulombe and Lée's ([21] and [47]) results, but provide additional insight into the driving forces behind economic growth. Where my results differ from Coulombe and Lee is in finding divergence in the cross-section distribution for GPP per hour worked. Coulombe and Lee [47] claim that the speed of convergence increases as disaggregation moves from per capita to per worker to per hour. Although my results show that dispersion may be falling over time, activity within the cross-section distribution may be diverging. GPP per hour tends to accumulate at opposite ends of the distribution.

F Quah and Sargent's Dynamic Index Model

The dynamic index model developed by Quah and Sargent (QS) [76] uses information from a wide range of cross-section data to explain aggregate fluctuations.

F.1 Theoretical Model

Let $\{y_{it}, i = 1, 2, \dots, N, t = 1, 2, \dots, T\}$ be an observed segment of a random field where the cross-section (N) and time series (T) dimensions are of the same order of magnitude. For example, y_i may be income per capita (relative to the national average) in province i . Suppose each y_i is affected by J common factors $u = (u_1, u_2, \dots, u_J)'$ and an idiosyncratic disturbance z_i . This can be written as,

$$(G.1) \quad y_{it} = \sum_{j=1}^J \alpha_{ij}(L) u_{jt} + z_{it}.$$

where u is a vector of orthogonal random walks; and z_i is zero mean, stationary, and has entries uncorrelated across all i as well as with the increments in u at all leads and lags. Observed y_{it} are therefore decomposed into first, factors common across all i and, second, factors specific to each i and orthogonal to all other i .

QS [76] assume the following structure for the impact of the common factors on y_i ,

$$(G.2) \quad \alpha_{ij}(L) = \sum_{m=0}^{M_a} a_{ij}(m) L^m, \quad M_a < \infty.$$

The u_j 's are integrated processes, with their first differences pairwise orthogonal. They have a finite autoregressive representation,

$$(G.3) \quad \gamma(L) \Delta u_t = \eta_{ut},$$

where $\gamma(L)$ is diagonal, with j -th entry given by,

$$(G.4) \quad 1 - g_j(1)L - g_j(2)L^2 - \dots - g_j(M_g)L^{M_g}, \quad M_g < \infty.$$

η_u is white noise, with mean 0 and the identity covariance matrix.

QS [76] further assume that each z_i is a finite order autogression,

$$(G.5) \beta_i(L)z_{it} = \epsilon_{it} \text{ for all } i$$

with

$$(G.6) \beta_i(L) = 1 - b_i(1)L - b_i(2)L^2 - \dots - b_i(M_b)L^{M_b}, \quad M_b < \infty.$$

Combining equations (F1) and (F5) and defining $\phi_{ij}(L) = \beta_i(L)\alpha_{ij}(L)$, produces,

$$(G.7) \beta_i(L)y_{it} = \sum_{j=1}^J \phi_{ij}(L)u_{jt} + \epsilon_{it}.$$

Exogenous variables, such as a constant or time trend—represented by $d_i w_t$ —can be added (without loss of generality) to (F7).

The model given by (F7) is translated into state space form, represented by a measurement equation (F8) and a transition equation (F9),⁹¹

$$(G.8) y_t = ax_t + dw_t + z_t,$$

and

$$(G.9) x_{t+1} = cx_t + \gamma_{t+1}.$$

The state vector x has dimension $O(N)$, which is ‘computationally intractable’ for large N and large T . The state-space representation given by equations (F8) and (F9) describes N time series in terms of an $O(N)$ -dimensional vector process.

However a large part of x —the idiosyncratic disturbances—is directly observable. The unobservable part—the common factors—has dimension $O(J)$, which is independent of N . QS [76] exploit this structure to give Kalman smoother projections and moment matrix calculations that are (effectively) independent of N .

The intuition is that increasing the cross-section dimension N can only help to estimate more precisely the state and its cross moments. This must imply the same property for estimates of the parameters. Deleting entries of y does not change the orthogonality properties on a reduced version of (F8) and (F9). Therefore—if the model is correctly specified—estimators that exploit the orthogonality conditions in (F8) and (F9) remain consistent independent of N . Smaller N -systems imply less efficient estimators.

When the unknown distribution of (y, u, z, w) generates a conditional expectation $E(u|y, z)$ that equals the linear projection of u on y and z , then standard Kalman smoother calculations yield the conditional expectations $E(x_t x_t' | y, w)$ and $E(x_t | y, w)$, taking as given a particular setting for the unknown parameters.

Iteration on this scheme is the EM algorithm.⁹² Each iteration of EM requires a Kalman filter and smoother followed by straightforward regression calculations. The algorithm consists of two steps which are iterated to convergence:

⁹¹State space representation is a means of summarizing finite processes. See QS [76] for details of how this is done.

⁹²The EM algorithm is a method for maximizing a likelihood function when there are missing observations.

an estimation and a maximization step. The maximization step calculates the maximum likelihood estimates of all unknown parameters conditional on a full data set. The estimation step constructs estimates of the sufficient statistics of the problem, conditional on the observed data and the parameters. Missing observations are estimated based on the parameter values at one step of the iteration and then the likelihood function is maximized assuming that this is the full observable data set in the other.⁹³ Under weak regularity conditions the EM algorithm is guaranteed to converge to a point that solves first order conditions.

F.2 Application

I investigate the extent to which two observable measures, aggregate employment and national GNP, can capture the cross-correlation across provincial GDP data (1967-1990), and compare this with one- and two-index representations. Figures F1 and F2 plot the sample standard deviations from second-order autoregressions over the period 1968-1988 (including a constant and time-trend). Figure F1 graphs the residual sample standard deviation when employment at lags -2 through 2 are included as additional regressors, versus that without. If points fall below the 45 degree line, comovements are better described by the model represented on the vertical axis than that on the horizontal. Total employment does appear to contain information on provincial GDP fluctuations. A similar result arises with national GNP data (Figure F2). Figure F3 compares these two observable measures—both appear to give similar descriptions of underlying comovements in the data. A one-index model provides little additional explanatory power over the observable measures.

Figures F4 and F5 plot standard deviations of innovations in the disturbances under a two-index model, against the residuals in province-by-province projections including total employment (Figure F4) and total GNP (Figure F5) respectively. The vertical axis describes the innovations upon removing two common unobservable indexes and imposing extensive orthogonality conditions; the horizontal axis describes the innovations upon removing total employment or total GNP respectively, without requiring the resulting innovations to be orthogonal across provinces. The results suggest that the two-index model provides a better description of the underlying common factors in provincial GDP than does either aggregate GNP or aggregate employment.

The two-index model provides more information about the cross-correlation in the data, but the combined effect of the two indices in the two-index model is very similar to the single-index model.⁹⁴

⁹³Watson and Engle [81] present the methodology for estimation using the EM algorithm in more detail. The estimation step gives parameter values at step $k + 1$ based on moment matrices estimated in step k . Each of the moment matrices can be constructed from the output of one pass through the Kalman smoother.

⁹⁴The correlation coefficient between the single index and the summation of the two indices in the two-index model is 0.996508.

Figure F1
Sample Standard Deviations:
Including Employment (1968-1988)

Figure F2
Sample Standard Deviations:
Including GNP (1968-1988)

Figure F3
Sample Standard Deviations:
Including Employment vs. Including GNP (1968-1988)

Figure F4
Sample Standard Deviations:
Two-Index Model vs. Total Employment (1968-1988)

Figure F5
Sample Standard Deviations:
Two-Index Model vs. Total GNP (1968-1988)