FACULTEIT ECONOMISCHE EN
TOEGEPASTE ECONOMISCHE
WETENSCHAPPEN

KATHOLIEKE
UNIVERSITEIT
LEUVEN

Essays on Endogeneity and Parameter Heterogeneity in Cross-section and Panel Data

Proefschrift voorgedragen
tot het behalen van de graad
van Doctor in de Economische
Wetenschappen
door
Rembert De Blander

Nummer 219 2006

Daar de proefschriften in de reeks van de Faculteit Economische en Toegepaste Economische Wetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

Contents

Contents i

Figures v

Tables vii

Dankwoord ix

Introduction 1

CRC Model and Returns to Schooling 2

Unit Root Tests in Panels 4

1 The CRC Model 7

1.1 Introduction 7

1.2 Returns to Schooling 9

1.3 Existing Estimators 12

1.4 Comparison 16

1.5 Conclusion 18

2 A Simple Estimator for the CRC Model 19

2.1 Introduction 19

2.2 Estimation 22

2.2.1 Derivation 22

2.2.2 Assumption B: Strength and Applicability 25

2.3 Comparison 26

2.4 Extension 28

2.5 Simulations 29

2.6 Conclusion 33

3 Semiparametric Estimation of the CRC Model **35**
3.1 Introduction . . . 35
3.2 Theoretical Framework . . . 37
3.3 Estimation . . . 39
3.4 Conclusion . . . 49

4 IV Estimation of the CRC Model in Panels **51**
4.1 Introduction . . . 51
4.2 Framework and Preliminaries . . . 53
4.3 Identification . . . 57
4.4 Estimation . . . 63
4.4.1 IV Estimation . . . 63
4.4.2 Aitken Estimation . . . 65
4.4.3 Covariance Estimation . . . 66
4.4.4 Feasible Aitken Estimation . . . 69
4.5 Concluding Remarks . . . 71

5 Unit Root Tests in Panels **73**
5.1 Introduction . . . 73
5.2 Theoretical Framework . . . 76
5.3 Test Statistics . . . 79
5.3.1 Model 1 . . . 79
5.3.2 Model 2 . . . 83
5.3.3 Model 3 . . . 88
5.4 Concluding Remarks . . . 92

A A Simple Estimator for the CRC Model **95**
A.1 Decomposition of $h_{\varepsilon;i}$ . . . 95
A.2 Inconsistency of the SB Estimator . . . 96
A.3 Asymptotic Normality . . . 99

B Semiparametric Estimation of the CRC Model **101**
B.1 Preliminaries . . . 101
B.2 Proofs of Theorems . . . 106

C IV Estimation of the CRC Model in Panels **113**
C.1 Identification 113
C.2 Estimation 116
C.2.1 IV Estimation 116
C.2.2 Inverse of Ω and Order of $X_1'\Omega^{-1}X_2$ 119
C.2.2.1 Inverse of Ω 119
C.2.2.2 Order of $X_1'\Omega^{-1}X_2$ 120
C.2.3 Aitken Estimation 123
C.2.4 Covariance Estimation 125
C.2.5 Feasible Aitken Estimation 127

D Unit Roots in Panels **139**
D.1 Preliminaries 139
D.2 Proofs for Model 1 142
D.3 Proofs for Model 2 144
D.4 Proof for Model 3 150

Bibliography **157**

List of Doctoral Dissertations **165**

List of Figures

5.1 The Inconsistency of $\hat{\varphi}_{LSP;HT}$ due to Neglecting Serial Correlation. 81
5.2 The Probability Limit of $\hat{\rho}_{LSDV}$ when $S = 1$. 86
5.3 The Probability Limit of $\hat{\rho}_{LSDT}$ when $S = 1$. 91

List of Tables

2.1 Estimation results: homoskedastic v_i, different specifications for $h_{\varepsilon;i}$, $h_{u;i} = v_i$. 30
2.2 Estimation results: heteroskedastic v_i, $h_{\varepsilon;i} = v_i$, $h_{u;i} = v_i$. 31
2.3 Estimation results: heteroskedastic v_i, $h_{\varepsilon;i} = v_i + v_i^2 - \mathrm{E}\left[v_i + v_i^2 \mid x_i, w_i, q_i\right]$, $h_{u;i} = v_i$. 31

5.1 The maxima of $\hat{\rho}_{LSDT}(\rho, T)$ for different values of T when $S = 1$. 92
5.2 The maxima of $\hat{\rho}_{LSDT}(\rho, T)$ for different values of T when $S = -\infty$. 93

Dankwoord

Zoals zovelen koester ik de geruststellende illusie dat eigen succes voornamelijk toe te schrijven is aan persoonlijke inspanningen, en dus niet meer dan verdiend is. Nochtans, maak ik hier graag gebruik van de mogelijkheid om het illusoir karakter van bovenstaande meritocratische gedachte te benadrukken en iedereen te bedanken zonder wie deze thesis ...

- Famke, voor het nalezen van eerdere versies, aanmoediging als het niet meer vlotte en liefde tout court;
- Andrée en Walter, voor nature én nurture;
- Lut en Frans, voor Famke;
- Katrien en Dirk, Helke Ruth en Stien, en Herman, voor de begrijpende oren;
- Iris en Ralph, Johan en Marina, Tom en de andere rock'n'rollers, voor de welkome dosissen fun;
- het HIVA en meer bepaald Ides Nicaise, voor de nodige flexibiliteit;
- Marno Verbeek, promotor en Geert Dhaene, copromotor;
- Hans De Wachter, Bertrand Melenberg, Frans Spinnewyn en Jean-Pierre Urbain, leden van de doctoraatscommissie.

Hartelijk dank aan allen voor hun respectievelijke bijdrage.

Introduction

In many applications, one starts from a microeconomic model where the parameters are assumed to be constant over all units of observation. Despite its limitations, this approach is nevertheless widely applied, mainly because of "its intuitive appeal, computational simplicity, and ease in interpretation of the result" (Akkina (1974)). Indeed, already half a century ago, it has been convincingly argued that in some situations, regression coefficients can reasonably be assumed to be "chance variables", instead of constants (Wald (1947)). In such a case, their estimated value can then be interpreted as the mean for all units or as the corresponding macro-parameter. This interpretation, while leading to large biases in the general case of a dynamic model (Pesaran and Smith (1995)), is correct in the restrictive case of a well-specified linear static model, as shown by Zellner (1969). Consequently there is no aggregation bias in the latter framework. As a result, a large body of literature has ensued on stochastic parameter regression or random coefficient estimation.

The combination of parameter heterogeneity with endogeneity is a relatively recent extension of the random coefficient model in which the question is raised how to estimate the population average of random coefficients correlated with the regressors. This so called correlated random coefficient (CRC) model is explicitly derived from economic theory by Card (2001) in the context of the returns to schooling, but is generally applicable where outcomes are influenced by choice variables. The main part of this thesis presents some estimators for CRC models.

The aggregation problem is also relevant to the next topic covered in this thesis, where I develop a panel data unit root test in the presence of heterogeneous constant terms and time trends. In contrast with previous tests, this new test combines a fixed time dimension with first order correlation in the error terms.

Below, I will demonstrate the empirical relevance of the CRC model by means of the returns to education as well as the different methods that have been used to estimate them. Finally, I will explain the notion of a unit root and illustrate the usefullness of testing for

it in applied economic work.

CRC Model and Returns to Schooling

As the relationship between education and socio-economic success is closely being scrutinized in a number of fields, the attention in economics is mainly focused on the monetary returns to schooling, which give important guidance regarding both private and public investments in education. Several factors can be identified that triggered a renewed interest in the subject: cost-benefit concerns about higher education due to higher entrance-rates, increasing attention to the role of human capital in economic growth and development and changes in the wage structure. The latter consists mainly of increasing wage inequality, largely due to the rising wage premium (Tobias (2003)). From the presumption that 'coerced' highly educated groups will benefit from identical high returns as the voluntary highly educated observed now, seems to stem a widespread belief that education is the key variable to counter a number of problems. These range from mitigating intergenerational correlations in income (Ashenfelter and Rouse (1998)) to effacing wage inequalities, whose increase is possibly related to globalization (Martins and Pereira (2004)). Whether increased education is such a panacea can only be assessed once the returns to education are correctly estimated.

The basic schooling model takes the logarithm of the wage to be linearly dependent on the number of completed years of education, a number of covariates and an error term (Mincer (1974)). The economic return to education is then equated to the estimated coefficient of the schooling variable.

Supposing for now that the parameters of the schooling model are correctly assumed to be constant, two sources of bias are traditionally associated with their ordinary least squares (OLS) estimation (Harmon and Walker (1995), Card (2001), Heckman and Vytlacil (2001)). First, failure to control for ability, which is positively correlated with both education and earnings, causes the OLS estimates for the return to schooling to be upwardly biased. Secondly, measurement error will bias OLS estimates downwardly, but will not be elaborated upon.

The ability bias reduces to an omitted variables problem (Heckman and Vytlacil (2001)), for which several approaches have been proposed. The evident solution is to include information like IQ or some other test score when available (Heckman and Vytlacil (2001)). A number of studies uses twin or sibling data to difference out unobserved ability (Ashen-

felter and Rouse (1998)). Another solution would be to treat ability as fixed over time and estimate the basic schooling model by the fixed effects estimator when panel data are available that contain individuals who return to school (Angrist and Newey (1991)). Some studies control for unobserved ability by means of selection models (Heckman, Tobias and Vytlacil (2001)). Instrumental variable (IV) estimation, finally, relies on the existence of instruments: exogenous variables that are correlated with schooling but uncorrelated with ability.

Instruments can originate from 'natural experiments', such as the raising of the minimum school-leaving age (Harmon and Walker (1995)), from more static features of the education system, such as the distance to the nearest college (Kane and Rouse (1995)), or from personal characteristics unrelated with ability, such as quarter of birth, which may affect the time spent in mandatory school attendance (Angrist and Krueger (1991)). Recently, these instruments, all weakly correlated with education, are being criticized as producing counter-intuitive results, since their estimate of the return to schooling lies above the OLS estimate. As a consequence, parental background variables come again into focus as instruments for schooling (Lauer (2003), Lemke and Rischall (2003), Agnarsson and Carlin (2002)).

There are several reasons why the assumption of fixed parameters might be incorrect. Individually differing returns to education can be caused on the one hand by factors known to the individual at decision time, but unobservable to the econometrician. On the other hand, they can be due to random shocks occurring after the decision is made. The variance of the random returns to schooling can also change over time due to supply and demand effects in certain segments of the labour market (Harmon, Hogan and Walker (2003)).

Assuming now that the parameters of the schooling model are random, introduces new potential sources of bias for the OLS estimates. Since the schooling choice is the result of a rational optimization, schooling and its return are expected to be positively correlated (Harmon and Walker (1995)). This situation corresponds to the definition of the correlated coefficient model.

Making abstraction of any observed covariates, the CRC-schooling model can be restated as a model in which the logarithm of the wage is linearly dependent on schooling, on the interaction between schooling and a first (multiplicative) error term and on a second (additive) error term. In this model both error terms are correlated with schooling and the return to schooling is fixed by construction. The question now naturally arises under which conditions the mean return to schooling can be consistently estimated by which estimator.

A first method estimates the return to schooling by IV, and makes the supplementary assumption that the interaction term is mean-independent of the instruments (Heckman and Vytlacil (1998), Wooldridge (2003a, 1997)). This latter condition is called the constancy of conditional covariances (CCC) condition. A second approach lies in the class of selection or control function models. Originally developed as an extension to continuous choice variables of Heckman's (1979, 1976) sample selection model, Garen's (1984) selectivity bias (SB) method first estimates schooling as a function of its instruments. Secondly, it assumes that both the multiplicative and additive error terms are linearly dependent (LD) on the schooling residual, which can be considered as a proxy for ability. Finally, 'ability' and its interaction with schooling are included in the wage equation, which is estimated by OLS. In a CRC context, the conditions for consistency imposed by the SB method are stronger than the ones imposed by IV. In Chapter 1 an overview of the existing estimators is presented.

My contribution toward developing estimators for CRC models is twofold: a first line of approach tries to adapt the SB method to weaker assumptions, a second one extends the IV method to panel data. In Chapter 2, I not only include 'ability' and its interaction with schooling, but also its square in the wage equation. This allows to weaken the LD condition to one comparable with CCC. Chapter 3 rids us of the LD as well as of the CCC condition, by estimating the wage equation semiparametrically, where the parametric part consists of schooling and the nonparametric part of ability and its interaction with schooling. In Chapter 4, finally, IV estimation of CRC models is extended to panel data by considering the use of instruments in the line of Hausman and Taylor (1981). Observing schooling for a number of individuals over time, it can be written as the sum of its mean, an individual-specific component, a period-specific component and an idiosyncratic part. If we are willing to make the assumption that ability is constant over time, the period-specific and idiosyncratic components of schooling can be used as instruments. Obviously, the CCC condition needs to be imposed for this IV estimator to be consistent.

Unit Root Tests in Panels

Economists face a lot of time series that are nonstationary. These are upwardly trending instead of fluctuating around some constant mean and can either be modeled by a deterministic time trend, or by a unit root process. The most important difference between both models is that for a trend-stationary process shocks wear off, while for an integrated

process they have an everlasting effect. To be able to distinguish between both models is important in a number of economic applications. For example, do economic recessions permanently depress the future output level, or are they merely temporary downturns and will the GNP catch up during recovery (Hamilton (1994))?

A related topic is the problem of cointegration, which occurs when a number of time series is integrated, but some linear combination of them is stationary. This means that a long-run equilibrium relation connects these time series, despite the fact that each individual one is permanently affected by shocks. Remark that in order to test for cointegration, we need to test for unit roots several times, once for each time series and once for the residuals of the linear combination. A hypothesis that can be tested by this framework is the purchasing power parity theory, which states that the effective price of goods, apart from transportation costs, is identical in different locations (Hamilton (1994)). Another testable hypothesis would be Gibrat's law, which states that growth rate and size (of firms, cities,...) are independent. If we autoregress the logarithm of size, a unit root would correspond to Gibrat's law (Resende (2004)).

The nature of the questions that can be answered by unit root tests only partly explains the attention they get. The importance of this topic is also explained by the fact that the asymptotics for estimated parameters of unit root processes are fundamentally different from those for stationary processes. Instead of being normal, they can be described in terms of functionals of Brownian motion.

The augmented Dickey-Fuller test and the semi-parametric Phillips-Perron test procedures are well-known tests for the presence of a unit root (Hamilton (1994)). Unfortunately, in finite samples these test procedures have limited power against near unit root alternatives. The extension of these techniques to panel data generates more powerful unit root tests, which in addition are shown to converge to a normal distribution, assuming data that are independently distributed over individuals (Quah (1984), Levin and Lin (1993), Im, Pesaran and Shin (1996) and Harris and Tzavalis (1999), Levin et al. (2002)).

Harris and Tzavalis (1999) develop asymptotic unit root tests for first-order autoregressive panel data models with fixed time dimension, based on the normalized least squares estimators of the autoregressive coefficients. They show that the limiting distribution of these test statistics is standard normal, and that the convergence rate is equal to $\sqrt{N}$, which is identical to the convergence rate for stationary panel data. When fixed effects or individual deterministic trends are included in the regression, the LS estimators are asymptotically downward biased, even as $T \to \infty$, which contrasts with the stationary

case.

In Chapter 5, I extend the last approach by allowing the error terms to be generated by an $AR(1)$ process. Since in most empirical studies the number of cross-sectional units is large relative to the number of time series observations, tests for fixed time dimension seem particularly useful.

Chapter 1

The Correlated Random Coefficient Model: Assumptions, Estimators and the Returns to Schooling Example

Abstract. This paper presents the correlated random coefficient (CRC) model. In addition, Card's (2001) demonstration how a theoretical model for schooling decisions results in a CRC model for returns to schooling, is recapitulated. Finally, an overview of existing estimators for the CRC model is given and the assumptions under which they are consistent are compared.
Keywords: Average treatment effect; Correlated random coefficient model; Unobserved heterogeneity
JEL classification: C21

1.1 Introduction of the CRC Model

Consider a model where the outcome for individual i is given by

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i'\alpha_{2,i} + \varepsilon_i, \tag{1.1}$$

with $i = 1, 2, \ldots, N$ and where the vector x_i has dimension K_1. The $L \times 1$ vector of treatment or choice variables, z_i, is assumed to have the following reduced form

$$z_i = \beta_0 + B_1 x_i + B_2 w_i + v_i, \tag{1.2}$$

with w_i a vector of dimension $K_2 \geq L$. This inequality is the usual identification restriction, which allows the use of the vector w_i as a set of instruments for z_i. The variables in (x_i, w_i) are strictly exogenous in a sense to be described later.

The $L \times 1$ vector of heterogeneous treatment effects $\alpha_{2,i}$ is a linear combination of the observed vector (x_i, w_i, q_i) and the unobserved heterogeneity, u_i,

$$\alpha_{2,i} = \gamma_0 + \Gamma_1 x_i + \Gamma_2 w_i + \Gamma_3 q_i + u_i, \tag{1.3}$$

where the vector q_i has dimension $K_3 \geq 0$. The parameters of interest are the average treatment effects (ATE)[1],

$$\mathrm{E}[\alpha_{2,i}] = \gamma_0 + \Gamma_1 \cdot \mathrm{E}[x_i] + \Gamma_2 \cdot \mathrm{E}[w_i] + \Gamma_3 \cdot \mathrm{E}[q_i], \tag{1.4}$$

which suggests to restate (1.3) as

$$\alpha_{2,i} = \gamma_0^* + \Gamma_1 \tilde{x}_i + \Gamma_2 \tilde{w}_i + \Gamma_3 \tilde{q}_i + u_i,$$

with $\gamma_0^* = \mathrm{E}[\alpha_{2,i}]$, $\tilde{x}_i = x_i - \mathrm{E}[x_i]$ and similarly for the other variables.

The model consisting of equations (1.1)-(1.3), combined with the following assumption, was first called the Correlated Random Coefficient (CRC) model by Heckman and Vytlacil (1998).

Assumption A. The observed variables $(x_i', w_i', q_i', y_i, z_i')$ are independently distributed across i, and the variables in (x_i', w_i', q_i') are strictly exogenous in the model (1.1)-(1.3), i.e.

$$\mathrm{E}[\varepsilon_i \mid x_i, w_i, q_i] = 0, \tag{1.5}$$
$$\mathrm{E}[u_i \mid x_i, w_i, q_i] = 0, \tag{1.6}$$
$$\mathrm{E}[v_i \mid x_i, w_i, q_i] = 0. \tag{1.7}$$

The crucial feature of the CRC model is the presence, in the outcome equation, of interaction terms between endogenous choice variables and unobserved heterogeneity. In the next section, a theoretical justification of the CRC model will be given in the context of the returns to schooling.

[1] Following Wooldridge (2003a).

1.2 Returns to Schooling as a CRC Model

In this section the schooling choice model outlined in Card (2001) is reviewed. Assuming an infinite planning horizon starting at the end of compulsory school attendance, and a once-and-for-all decision on when to enter the workforce at that moment ($t = 0$), the lifecycle utility conditional on the chosen years of post-compulsory schooling (z) and a given consumption profile ($c(t)$), is given by

$$V(z, c(\cdot)) = \int_0^z \{u[c(t)] - \phi(t)\} \exp[-\rho t]\, dt + \int_z^\infty u[c(t)] \exp[-\rho t]\, dt, \tag{1.8}$$

where $u(\cdot)$ is the instantaneous utility of consumption, $\phi(t)$ is the relative disutility of being a student versus being a member of the workforce, and ρ is the subjective discount rate.

Making abstraction of tuition fees and the possibility to work part time while being a student, and assuming free borrowing and lending at the fixed interest rate R, the intertemporal budget constraint is given by

$$\int_0^\infty c(t) \exp[-Rt]\, dt = \int_z^\infty y(z, t) \exp[-Rt]\, dt, \tag{1.9}$$

where $y(z, t)$ is the amount of earnings at time t when z years of non-compulsory schooling are completed. Assuming that earnings are multiplicatively separable in a function of post-compulsory schooling years and a function of post-schooling experience, as in

$$y(z, t) = f(z)\, g(t - z), \qquad g(0) = 1,$$

maximization of $V(z, c(\cdot))$ subject to the budget constraint (1.9), is equivalent to maximizing

$$\begin{aligned} &\int_0^z \{u[c(t)] - \phi(t)\} \exp[-\rho t]\, dt + \int_z^\infty u[c(t)] \exp[-\rho t]\, dt \\ + \quad &\lambda \exp[-RS] \int_0^\infty f(z)\, g(\tau) \exp[-R\tau]\, d\tau - \lambda \int_0^\infty c(t) \exp[-Rt]\, dt, \end{aligned}$$

where λ is the Lagrange multiplier of the constrained optimization. Taking the derivative with respect to z results in

$$-\phi(z) \exp[-\rho z] + \lambda \exp[-Rz] \{f'(z) - Rf(z)\} h(R),$$

where

$$h(R) = \int_0^\infty g(\tau) \exp[-R\tau] \, d\tau.$$

As a consequence we have that the first-order condition with respect to z can be written as[2]

$$\lambda \exp[-Rz] \{MB(z) - MC(z)\} = 0,$$

where the marginal benefit of the zth year of schooling arises from the increase $f'(z) h(R)$ in lifecycle wealth expressed in period z currency. The marginal cost consists of the prolonged disutility of schooling, $\phi(z) \exp[-(\rho - R) z] / \lambda$, plus the monetary loss of postponing the onset of wages with one period, $Rf(z) h(R)$. The first-order condition with respect to z can also be written in the form

$$\frac{f'(z)}{f(z)} = R + \frac{\exp[-(\rho - R) z] \phi(z)}{\lambda f(z) h(R)}. \tag{1.10}$$

Introducing a last assumption that the instantaneous utility of consumption is represented by a logarithmic function

$$u[c(t)] = \log[c(t)],$$

the first-order condition for an optimal consumption profile can be restated as

$$\int_0^\infty \frac{\exp[-\rho t]}{c(t)} dt = \lambda \int_0^\infty \exp[-Rt] \, dt,$$

which has $c(t) = \exp[-(\rho - R) t] / \lambda$ as a solution. Combined with the budget constraint (1.9) this implies the relationship

$$\frac{1}{\lambda} = \rho \exp[-Rz] f(z) h(R),$$

which further modifies the first-order condition (1.10) to

[2] Assuming $\frac{\partial MC(z)}{\partial z} > \frac{\partial MB(z)}{\partial z}$.

$$\frac{f'(z)}{f(z)} = R + \rho \exp[-\rho z]\,\phi(z) \equiv d(z). \tag{1.11}$$

Equation (1.11) states that the optimal choice is characterized by the equality of marginal rate of return to schooling, $f'(z)/f(z)$, and marginal cost of schooling, $d(z)$. Heterogeneity in the optimal schooling choice can arise from individual differences in either side of equation (1.11). To make this model operational, both sides of equation (1.11) are assumed to be linear in z with individual-specific intercepts, and homogenous slopes

$$\frac{f'(z)}{f(z)} = b_i - k_1 z \qquad k_1 \geq 0, \tag{1.12}$$

$$d(z) = d_i + k_2 z \qquad k_2 > 0. \tag{1.13}$$

Here (b_i, d_i) is a vector of random variables with mean (b, d). Solving for the optimal level of schooling yields

$$z_i^* = \frac{b_i - d_i}{k}, \tag{1.14}$$

where $k = k_1 + k_2 > 0$. The sources of heterogeneity in this model are b_i: individuals have different returns to schooling, which corresponds loosely to differences in ability; and d_i: individuals have unequal rates of substitution between schooling and future earnings (Card (1995)), which corresponds to variation in opportunity costs or in access to funds.

At the optimal schooling level (z_i^*) as defined in (1.14), the individual's marginal return to schooling (1.12) becomes

$$b_i - k_1 z_i^* = b_i \left(1 - \frac{k_1}{k}\right) + d_i \frac{k_1}{k},$$

which gives rise to a nondegenerate distribution except in some very restrictive and implausible cases. Integration of equation (1.12) leads to the following structural model of earnings

$$\log I_i = a_i + b_i z_i - \frac{k_1}{2} z_i^2. \tag{1.15}$$

In this notation $\log I_i$ corresponds to y_i in the previous section.

This simple model illustrates two things. First of all, the return to schooling is in general not independent of the level of schooling. It can be either functionally dependent on z_i if there is a diminishing return to schooling ($k_1 > 0$), or it can be stochastically

dependent on schooling through any stochastic dependence between schooling and ability (b_i). Furthermore, although a_i in equation (1.15) is not functionally dependent on z_i, it can be stochastically dependent on (b_i, d_i), which in turn questions the exogeneity of z_i. These two conclusions theoretically underpin the adoption of the CRC model. In the next section I will review the estimators for the CRC model that are described in the literature.

1.3 Existing Estimators

Returning to model (1.1)-(1.3) and plugging (1.3) into (1.1), the outcome equation can be rewritten as

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i^{*\prime}\gamma + z_i'u_i + \varepsilon_i. \tag{1.16}$$

where[3] $\gamma = \text{Vec}\left((\gamma_0, \Gamma_1, \Gamma_2, \Gamma_3)'\right)$ and $z_i^* = \text{Vec}\left((1, x_i', w_i', q_i')' z_i'\right)$.[4]

Garen (1984) considers a single choice variable ($L = 1$) and assumes

$$\begin{pmatrix} \varepsilon_i \\ u_i \\ v_i \end{pmatrix} \sim NID\left(0; \Sigma\right), \tag{1.17}$$

which allows him to rewrite (1.16) as

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i^{*\prime}\gamma + v_i\alpha_3 + z_i v_i \alpha_4 + z_i\tilde{u}_i + \tilde{\varepsilon}_i,$$

with $\text{E}\left[(\tilde{\varepsilon}_i, \tilde{u}_i) \mid x_i, w_i, q_i, v_i\right] = 0$. He then proposes to estimate $\alpha = (\alpha_0, \alpha_1', \gamma', \alpha_3, \alpha_4)'$ by OLS,

$$\hat{\alpha}_{SB} = \left(\sum_{i=1}^{N} s_i s_i'\right)^{-1} \left(\sum_{i=1}^{N} s_i y_i\right), \tag{1.18}$$

where $s_i = (1, x_i', z_i^{*\prime}, z_i\hat{v}_i, \hat{v}_i)'$ and $\hat{v}_i$ the least squares residual from (1.2), and terms this

[3] The matrix operation $\text{Vec}\,(X)$ creates a $JK \times 1$ column vector by stacking the columns of the $J \times K$ matrix X on top of each other. Likewise, $\text{Vech}\,(Y)$ creates a $J\,(J+1)\,/2 \times 1$ column vector by columnwise stacking the lower triangular part of the $J \times J$ matrix Y, starting at the diagonal elements.

[4] Taking into account the suggestion following (1.4), we could redefine γ and z_i as $\gamma = \text{Vec}\left((\gamma_0^*, \Gamma_1, \Gamma_2, \Gamma_3)'\right)$, respectively $z_i^* = \text{Vec}\left((1, \tilde{w}_i', \tilde{x}_i', \tilde{q}_i')' z_i'\right)$, in which case the ATE is immediately given as γ_0^*.

procedure the selectivity bias (SB) method.

Wooldridge (1997) also considers a single choice variable ($L = 1$), but assumes a simplified parameter equation (1.3), with $(\Gamma_1, \Gamma_2, \Gamma_3) = 0$. In addition to Assumption A, he makes the assumptions

$$\begin{aligned} \mathrm{E}\left[u_i \mid x_i, w_i, v_i\right] &= v_i \alpha_4 && (1.19) \\ \mathrm{E}\left[v_i^2 \mid x_i, w_i\right] &= \mathrm{E}\left[v_i^2\right], && (1.20) \end{aligned}$$

under which he considers the standard 2SLS estimator

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\gamma}_0 \end{pmatrix}_{IV} = \left(\sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ w_i \end{pmatrix} (1, x_i', z_i') \right)^{-1} \left(\sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ w_i \end{pmatrix} y_i \right).$$

He then shows that $(\hat{\alpha}_1', \hat{\gamma}_0)_{IV}$ is consistent for (α_1', γ_0), where γ_0 is equal to the ATE, $\mathrm{E}\left[\alpha_{2,i}\right]$, as defined in (1.4), in the special case that $(\Gamma_1, \Gamma_2, \Gamma_3) = 0$. Remark that $\hat{\alpha}_0$ is inconsistent for the constant term α_0.[5]

Heckman and Vytlacil (1998) estimate the CRC model (1.1)-(1.3) using the two-step plug-in estimator

$$\begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\gamma} \end{pmatrix}_{2S} = \left(\sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ \hat{z}_i^* \end{pmatrix} (1, x_i', \hat{z}_i^{*\prime}) \right)^{-1} \left(\sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ \hat{z}_i^* \end{pmatrix} y_i \right),$$

with $\hat{z}_i^* = \mathrm{Vec}\left((1, x_i', w_i', q_i')' \hat{z}_i'\right)$ and $\hat{z}_i$ the least squares predictor of (1.2), under Assumption A and the assumption that

$$\mathrm{E}\left[v_i u_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[v_i u_i\right], \qquad (1.21)$$

which they term conditional homoskedasticity of covariances, but which will here be called constancy of conditional covariances. Under (1.21) the constant term is neither identified. Identification of $(\hat{\alpha}_1', \hat{\gamma}_0)_{2S}$ requires the covariance matrix of $(1, x_i', \hat{z}_i^{*\prime})$ to be of rank $1 + K_1 + L\left(1 + K_1 + K_2 + K_3\right)$.

Wooldridge (2003a) does not make use of the linear specification (1.2), but simply

[5] Wooldridge also proposes an alternative set of assumptions under which it holds that $\mathrm{E}\left[v_i \mid x_i, w_i, q_i, u_i\right] = u_i \zeta$ and $\mathrm{E}\left[u_i^2 \mid x_i, w_i, q_i\right] = \mathrm{E}\left[u_i^2\right]$.

assumes the existence of $K_2 \geq L$ instrumental variables w_i, and weakens assumption (1.21) to

$$\mathrm{E}\left[z_i u_i \mid x_i, w_i, q_i\right] \;=\; \mathrm{E}\left[z_i u_i\right]. \tag{1.22}$$

Furthermore, he strengthens (1.3) with $\Gamma_2 = 0$ and $\Gamma_3 = 0$ to

$$\alpha_{2,i} \;=\; \gamma_0 + \Gamma_1 x_i + u_i, \tag{1.23}$$

i.e. he does not allow that the correlated coefficients depend on the instruments w_i nor on covariates not present in the outcome equation. Under these conditions, and the standard rank condition

$$\mathrm{Cov}\left[\left(1, x_i', w_i', \mathrm{Vech}\left(x_i x_i'\right)', \mathrm{Vec}\left(x_i w_i'\right)'\right)\right] \;=\; 1 + K_1 + L\left(1 + K_1\right),$$

he applies IV estimation to the outcome equation (1.16), using

$$\left(1, x_i', w_i', \mathrm{Vech}\left(x_i x_i'\right)', \mathrm{Vec}\left(x_i w_i'\right)'\right)$$

as instruments, which results in a consistent estimate of the parameter vector (α_1', γ'), but not of the constant term α_0 of (1.16). Note that the order condition for identification

$$\begin{aligned} \left(1 + K_1\right) L &\leq K_2 + K_1\left(\frac{\left(1 + K_1\right)}{2} + K_2\right) \\ &\Updownarrow \\ L &\leq K_2 + \frac{K_1}{2} \end{aligned}$$

holds under $K_2 \geq L$, which was assumed in the introductory section 1.1.

Combining the results of Heckman and Vytlacil (1998) and Wooldridge (2003a), consider now IV estimation of (1.16) under Wooldridge's (2003a) conditions, but without the restrictions $\Gamma_2 = 0$ and $\Gamma_3 = 0$, using

$$\left(1, x_i', w_i', \mathrm{Vech}\left(\left(x_i', w_i'\right)'\left(x_i', w_i'\right)\right)', \mathrm{Vec}\left(q_i\left(x_i', w_i'\right)\right)'\right) \tag{1.24}$$

as instruments. Again this IV estimator is consistent for the parameter vector (α_1', γ'), but

not for α_0. The standard rank condition becomes

$$\begin{aligned} & \text{Cov}\left[\left(1, x_i', w_i', \text{Vech}\left(\left(x_i', w_i'\right)'\left(x_i', w_i'\right)\right)', \text{Vec}\left(q_i\left(x_i', w_i'\right)\right)'\right)\right] \\ = \;& 1 + K_1 + L\left(1 + K_1 + K_2 + K_3\right) \end{aligned}$$

and the order condition

$$L\left(1 + K_1 + K_2 + K_3\right) \;\leq\; K_2 + \left(K_1 + K_2\right)\left(\frac{\left(K_1 + K_2 + 1\right)}{2} + K_3\right), \tag{1.25}$$

is no longer implied by $K_2 \geq L$, which was required below (1.2). Condition (1.25) is most binding for $K_3 = 0$, since an increase of K_3 by 1 adds L to the left-hand-side and $K_1 + K_2$ to the right-hand side, and $K_2 \geq L$ clearly implies $L \leq K_1 + K_2$. Consequently we require that

$$L\left(1 + K_1 + K_2\right) \;\leq\; K_2 + \left(K_1 + K_2\right)\frac{\left(K_1 + K_2 + 1\right)}{2},$$

which is most binding for $K_1 = 0$, since an increase of K_1 by 1 adds L to the left-hand-side and $1 + K_1 + K_2$ to the right-hand side. As a result, when $\Gamma_1 = 0$ and $\Gamma_3 = 0$, condition

$$L \;\leq\; \frac{K_2\left(K_2 + 3\right)}{2\left(K_2 + 1\right)}$$

is sufficient for (1.25), i.e without extra restrictions on Γ_2, we need twice the number of instruments, compared with standard IV. Finally, when $\Gamma_1 = 0$ and $\Gamma_2 = 0$, the usual identification restriction, $K_2 \geq L$, is sufficient for the order condition

$$L\left(1 + K_3\right) \;\leq\; K_2 + \left(K_1 + K_2\right)K_3.$$

Similarly, under Wooldrige's (1997) assumptions (1.19)-(1.20), but without $(\Gamma_1, \Gamma_2, \Gamma_3) = 0$, IV estimation of (1.16), using (1.24) as instruments, is consistent for the parameter vector (α_1', γ'), but not for α_0.

1.4 Comparison of Estimators[6]

Wooldridge's (1997) linearity (1.19) and homoskedasticity (1.20) assumptions are clearly weaker than Garen's (1984) normality assumption (1.17), since joint homoskedastic normality of u_i and v_i (1.17) implies linearity of the conditional mean (1.19) and homoskedasticity of the marginal variance (1.20). Assumptions (1.19)-(1.20), in turn, imply

$$\begin{aligned}
\mathrm{E}\left[v_i u_i \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[v_i \mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right] \mid x_i, w_i, q_i\right] \\
&= \mathrm{E}\left[v_i^2 \mid x_i, w_i, q_i\right] \alpha_4 \\
&= \mathrm{E}\left[v_i^2\right] \alpha_4 \\
&= \mathrm{E}\left[v_i \mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right]\right] \\
&= \mathrm{E}\left[v_i u_i\right],
\end{aligned}$$

i.e. constancy of conditional covariances (1.21) holds. The reverse, however, does not hold. Assumption A in combination with (1.21), finally, implies

$$\begin{aligned}
\mathrm{E}\left[z_i u_i \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[\left(\beta_0 + B_1 x_i + B_2 w_i + v_i\right) u_i \mid x_i, w_i, q_i\right] \\
&= \left(\beta_0 + B_1 x_i + B_2 w_i\right) \mathrm{E}\left[u_i \mid x_i, w_i, q_i\right] + \mathrm{E}\left[v_i u_i \mid x_i, w_i, q_i\right] \\
&= \mathrm{E}\left[v_i u_i\right] \\
&= \mathrm{E}\left[\left(\beta_0 + B_1 x_i + B_2 w_i\right) u_i\right] + \mathrm{E}\left[v_i u_i\right] \\
&= \mathrm{E}\left[z_i u_i\right],
\end{aligned}$$

i.e. (1.22) holds. This proves the following ordering from weak to strong:

1. Wooldridge's (2003a) conditional homoskedasticity of covariances (1.22),
2. Heckman and Vytlacil's (1998) conditional homoskedasticity of covariances (1.21),
3. Wooldridge's (1997) linearity (1.19) and homoskedasticity (1.20) assumptions,
4. Garen's (1984) normality assumption (1.17).

Thus Wooldridge's (2003a) assumptions, without the restrictions $\Gamma_2 = 0$ and $\Gamma_3 = 0$, are the weakest of the four sets of assumptions, sufficient for consistency of the respective estimators of the ATE (1.4).

[6] To simplify notation, I only consider models with a single choice variable ($L = 1$) in this section.

In this context it is useful to notice that assumption (1.17) can be relaxed to

$$\begin{aligned} \mathrm{E}\left[\varepsilon_i \mid x_i, w_i, q_i, v_i\right] &= v_i\alpha_3 \qquad (1.26) \\ \mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right] &= v_i\alpha_4, \qquad (1.27) \end{aligned}$$

while still maintaining consistency of Garen's (1984) SB estimator (1.18). Furthermore, no constancy of conditional covariances is required for its consistency.

While Wooldridge's (1.22) is weaker than (1.17), it is neither weaker nor stronger than Assumption A in combination with (1.26)-(1.27). To show this, we disprove that (1.22) implies (1.26)-(1.27) and vice versa, by counterexamples. Consider the case where (1.1)-(1.3), Assumption A and (1.26)-(1.27) are fulfilled, but where $\mathrm{E}\left[v_i^2 \mid x_i, w_i, q_i\right] = f\left(x_i, w_i, q_i\right)$. Under these circumstances it holds that

$$\begin{aligned} \mathrm{E}\left[z_i u_i \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[z_i \mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right] \mid x_i, w_i, q_i\right] \\ &= \mathrm{E}\left[\beta_0 v_i\alpha_4 + B_1 x_i v_i\alpha_4 + B_2 w_i v_i\alpha_4 + v_i^2\alpha_4 \mid x_i, w_i, q_i\right] \\ &= f\left(x_i, w_i, q_i\right)\alpha_4 \\ &\neq \mathrm{E}\left[z_i u_i\right], \end{aligned}$$

i.e. (1.22) does not hold. Conversely, the case where (1.22) holds, but where $\mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right] = g\left(v_i\right)$, with $g\left(\cdot\right)$ nonlinear, disproves that (1.22) implies (1.26)-(1.27).

The main advantages of Wooldridge's (2003a) IV estimator are that we do not need

- knowledge of the reduced form for z_i, nor
- the linearity assumptions (1.26) and (1.27).

The main advantage of Garen's (1984) SB estimator is that we do not need the constancy of conditional covariances assumption (1.22).

We can now ask the question which estimator is best suited to estimate the average influence of schooling on earnings, when earnings are specified as in (1.15):

$$y_i = \log I_i \;=\; a_i + b_i z_i - \frac{k_1}{2} z_i^2 + \varepsilon_i.$$

The ATE of schooling is here given by $\mathrm{E}\left[b_i - \frac{k_1}{2} z_i\right]$. Consider now the special case when $b_i = 0$.[7] I also assume that a_i can be written as $\alpha_0 + x_i'\alpha_1$. Under these circumstances,

[7] When $k_1 = 0$, we have the general case which was discussed above.

Wooldridge's (2003a) assumption (1.23), i.e. the correlated coefficients do not depend on the instruments w_i, is violated and his IV estimator can not be applied. However, under the assumptions made by Heckman and Vytlacil (1998), i.e. (1.1)-(1.3), Assumption A and (1.21), we have knowledge of the reduced form (1.2)

$$z_i = \beta_0 + B_1 x_i + B_2 w_i + v_i.$$

We can apply IV to (1.15) with (1.2) plugged in:

$$y_i = \alpha_0 + x_i'\alpha_1 - \frac{k_1}{2}\beta_0 z_i - \frac{k_1}{2}B_1 x_i z_i - \frac{k_1}{2}B_2 w_i z_i - \frac{k_1}{2}v_i z_i + \varepsilon_i,$$

using $\left(\text{Vech}\left((x_i', w_i')'\,(x_i', w_i')\right)'\right)$ as instruments. This results in a consistent estimate of the parameter vector, except the constant term. Remark that condition (1.21) modifies here to

$$\text{E}\left[v_i^2 \mid x_i, w_i, q_i\right] = \text{E}\left[v_i^2\right],$$

i.e. the reduced form (1.2) for z_i is assumed to be homoskedastic. Garen's (1984) SB method would consist of applying OLS to

$$y_i = \alpha_0 + x_i'\alpha_1 - \frac{k_1}{2}\beta_0 z_i - \frac{k_1}{2}B_1 x_i z_i - \frac{k_1}{2}B_2 w_i z_i - \frac{k_1}{2}z_i\hat{v}_i + \hat{v}_i\alpha_4 + \tilde{\varepsilon}_i,$$

and would not require homoskedasticity of z_i. Remark that only the linearity assumption (1.26) is needed here. (1.27) is automatically fulfilled since we have that $u_i = v_i$.

1.5 Conclusion

In this chapter the CRC model was presented and justified in the context of returns to schooling, using Card's (2001) schooling decision model. Some estimators for the CRC were reviewed and the conditions under which they are consistent, were compared. No clear-cut ordering in terms of strength of their assumptions was obtained.

Chapter 2

A Simple Estimator for the Correlated Random Coefficient Model under Weak Assumptions

Abstract. This paper presents an estimator for the Correlated Random Coefficient (CRC) model which is an extension of Garen's (1984) Selectivity Bias Method, compared to which the outcome equation is augmented with more terms.
Keywords: Average treatment effect; Correlated random coefficient model; Unobserved heterogeneity
JEL classification: C21

2.1 Introduction

Consider a model where the outcome for individual i is given by

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i'\alpha_{2,i} + \varepsilon_i, \tag{2.1}$$

with $i = 1, 2, \ldots, N$ and where the vector x_i has dimension K_1. The $L \times 1$ vector of treatment or choice variables, z_i, is assumed to have the following reduced form

$$z_i = \beta_0 + B_1 x_i + B_2 w_i + v_i, \tag{2.2}$$

with w_i a vector of dimension $K_2 \geq L$. This inequality is the usual identification restriction, which allows the use of the vector w_i as a set of instruments for z_i. The variables in (x_i', w_i') are strictly exogenous in a sense to be described later (see Assumption A below).

The vector of heterogeneous treatment effects $\alpha_{2,i}$ is a linear combination of the observed vector (x_i', w_i', q_i') and the unobserved heterogeneity, u_i

$$\alpha_{2,i} = \gamma_0 + \Gamma_1 x_i + \Gamma_2 w_i + \Gamma_3 q_i + u_i, \tag{2.3}$$

where the vector q_i has dimension $K_3 \geq 0$. The parameters of interest are the average treatment effects (ATE),

$$\mathrm{E}[\alpha_{2,i}] = \gamma_0 + \Gamma_1 \mathrm{E}[x_i] + \Gamma_2 \mathrm{E}[w_i] + \Gamma_3 \mathrm{E}[q_i],$$

which suggests to restate treatment effects $\alpha_{2,i}$ as $\gamma_0^* + \Gamma_1 \tilde{x}_i + \Gamma_2 \tilde{w}_i + \Gamma_3 \tilde{q}_i + u_i$, with $\gamma_0^* = \mathrm{E}[\alpha_{2,i}]$, $\tilde{x}_i = x_i - \mathrm{E}[x_i]$ and similarly for the other variables.

The model consisting of equations (2.1)-(2.3), combined with the following assumption, was first called the Correlated Random Coefficient (CRC) model by Heckman and Vytlacil (1998).

Assumption A. The observed variables $(x_i', w_i', q_i', y_i, z_i')$ are independently distributed across i, and the variables in (x_i', w_i', q_i') are strictly exogenous in the model (2.1)-(2.3), i.e.

$$\begin{aligned} \mathrm{E}[\varepsilon_i \mid x_i, w_i, q_i] &= 0, \\ \mathrm{E}[u_i \mid x_i, w_i, q_i] &= 0, \\ \mathrm{E}[v_i \mid x_i, w_i, q_i] &= 0. \end{aligned}$$

The crucial feature of the CRC model is the presence, in the outcome equation, of interaction terms between endogenous choice variables and unobserved heterogeneity. Card (2001) gives a theoretical justification of the CRC model in the context of the returns to schooling.

Garen (1984) assumes a single choice variable ($L = 1$) and joint normality of $(\varepsilon_i, u_i, v_i)$ and proposes the selectivity bias (SB) method, in which he estimates the outcome equation, expanded with both the residual of the choice equation ($\hat{v}_i$) and its cross-product with the treatment variable ($z_i \hat{v}_i$), by OLS. It can easily be shown, however, that by replacing the

joint normality assumption with the weaker linearity assumptions

$$\mathrm{E}\left[\varepsilon_i \mid x_i, w_i, q_i, v_i\right] = \alpha_3 v_i, \tag{2.4}$$

$$\mathrm{E}\left[u_i \mid x_i, w_i, q_i, v_i\right] = \alpha_4 v_i, \tag{2.5}$$

the SB method remains consistent.

In the special case that $(\Gamma_1, \Gamma_2, \Gamma_3) = 0$, Wooldridge (1997) shows that standard 2SLS of y_i on $(1, x_i, z_i)$, using $(1, x_i, w_i)$ as instruments, is consistent for the ATE under weaker assumptions than Garen's normality assumptions. Heckman and Vytlacil (1998) consider the CRC model (2.1)-(2.3) and derive a two-step plug-in estimator. While Wooldridge (1997) assumes that the relationship between the error term of the choice equation and the unobserved heterogeneity is linear, no such assumption is made by Heckman and Vytlacil (1998). Furthermore, Wooldridge's (1997) assumption of homoskedasticity is replaced with the weaker assumption of constancy of conditional covariances, $\mathrm{E}\left[v_i' u_i \mid x_i, w_i, q_i\right]$. Finally, assuming $(\Gamma_2, \Gamma_3) = 0$, Wooldridge (2003) applies IV estimation to the outcome equation (2.1) expanded with the interaction terms $\mathrm{Vec}\left(x_i z_i'\right)$, using $\left(1, x_i', w_i', \mathrm{Vech}\left(x_i x_i'\right)', \mathrm{Vec}\left(x_i w_i'\right)'\right)$ as instruments[1]. In addition, he requires a weaker version of the constancy of conditional covariances condition, $\mathrm{E}\left[z_i' v_i \mid x_i, w_i, q_i\right]$.

The aim of this chapter is to extend Garen's (1984) SB method by weakening the linearity assumptions (2.4)-(2.5), as well as by allowing for multiple treatments. To that end, I propose to expand the set of regressors with more elements compared to the SB method. The resulting outcome equation is subsequently estimated by OLS. This new estimator is consistent under conditions comparable to those required for the mentioned IV and plug-in estimators, while having a smaller variance, and allowing for a test of the possible correlation between treatment and its effect.

In the next section a new estimator is derived, which is compared with existing estimators in section 3. I consider an extension in section 4, while section 5 presents the results of a small Monte Carlo simulation and discusses some practical issues. Section 6, finally, concludes.

[1] The matrix operation $\mathrm{Vec}\left(X\right)$ creates a $JK \times 1$ column vector by stacking the columns of the $J \times K$ matrix X on top of each other. Likewise, $\mathrm{Vech}\left(Y\right)$ creates a $J\left(J+1\right)/2 \times 1$ column vector by columnwise stacking the lower triangular part of the $J \times J$ matrix Y, starting at the diagonal elements.

2.2 Estimation under Weak Assumptions

When the linearity assumptions (2.4)-(2.5) are dropped, including $\hat{v}_i$ and $z_i\hat{v}_i$ as extra regressors in outcome equation (2.1) does not totally remove the correlation between the regressors and the error term. Complete removal of aforementioned correlation is obtained, firstly, by including more parts of the error term ε_i in the set of regressors and, secondly, by an assumption comparable to the constancy of covariances (Heckman and Vytlacil (1998), Wooldridge (2003)).

2.2.1 Derivation

The model consisting of (2.1)-(2.3), is summarized as

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i^{*\prime}\gamma + z_i'u_i + \varepsilon_i, \tag{2.6}$$

where[2] $\gamma = \text{Vec}\left((\gamma_0, \Gamma_1, \Gamma_2, \Gamma_3)'\right)$ and $z_i^* = \text{Vec}\left((1, x_i', w_i', q_i')'\, z_i'\right)$. Define $h_{\varepsilon;i}$ and $h_{u;i}$ as the conditional expectations

$$\begin{aligned} h_{\varepsilon;i} &\equiv h_\varepsilon(x_i, w_i, q_i, v_i) = \text{E}\left[\varepsilon_i \mid x_i, w_i, q_i, v_i\right], \\ h_{u;i} &\equiv h_u(x_i, w_i, q_i, v_i) = \text{E}\left[u_i \mid x_i, w_i, q_i, v_i\right], \end{aligned} \tag{2.7}$$

and expression (2.6) can be written as

$$y_i = \alpha_0 + x_i'\alpha_1 + z_i^{*\prime}\gamma + z_i'h_{u;i} + h_{\varepsilon;i} + z_i'\tilde{u}_i + \tilde{\varepsilon}_i, \tag{2.8}$$

where

$$\begin{aligned} \text{E}\left[\tilde{\varepsilon}_i \mid x_i, w_i, q_i, v_i\right] &= 0 \\ \text{E}\left[\tilde{u}_i \mid x_i, w_i, q_i, v_i\right] &= 0. \end{aligned} \tag{2.9}$$

Inclusion of only v_i in the set of regressors[3] is equivalent to IV and would require the constancy of conditional covariances assumption (Heckman and Vytlacil (1998), Wooldridge (2003)). Under the general specification (2.7) instead of (2.4)-(2.5), Garen's (1984) SB

[2] Taking into account the suggestion following (2.3), we could redefine γ and z_i as $\gamma = \text{Vec}\left((\gamma_0^*, \Gamma_1, \Gamma_2, \Gamma_3)'\right)$, respectively $z_i^* = \text{Vec}\left((1, \tilde{w}_i', \tilde{x}_i', \tilde{q}_i')'\, z_i'\right)$, in which case the ATE is immediately given as γ_0^*.

[3] Next to $(1, x_i', z_i^{*\prime})$.

estimator[4] would be inconsistent[5], since both the regressors and the error term would contain components in v_i^2. The latter correlation between disturbances and regressors can be removed in the following way.

Define the $(K_1 + K_2 + K_3 + 1)(K_1 + K_2 + K_3 + 2)/2 \times 1$ vector containing the cross-products of all exogenous variables $r_i = \text{Vech}\left((1, x_i', w_i', q_i')'(1, x_i', w_i', q_i')\right)'$. The $L(L+1)/2 \times 1$ vector e_i is defined as the vector of crossmultiplied disturbances of (2.2), orthogonalized with respect to r_i

$$e_i = \text{Vech}(v_i v_i') - \text{E}\left[\text{Vech}(v_i v_i') r_i'\right] \left(\text{E}\left[r_i r_i'\right]\right)^{-1} r_i. \tag{2.10}$$

It is shown in Appendix A that

$$\begin{aligned} h_{\varepsilon;i} &= v_i'\theta_{\varepsilon 1} + \text{Vec}(x_i v_i')'\theta_{\varepsilon 2} + \text{Vec}(w_i v_i')'\theta_{\varepsilon 3} \\ &\quad + \text{Vec}(q_i v_i')'\theta_{\varepsilon 4} + e_i'\theta_{\varepsilon 5} + \mathcal{R}_{\varepsilon;i}. \end{aligned} \tag{2.11}$$

Remark that terms of degree zero in v_i are lacking in this decomposition. The remainder $\mathcal{R}_{\varepsilon;i}$ in (2.11) is uncorrelated with $(x_i', z_i^{*\prime}, v_i^{*\prime}, z_i' v_i)$ where

$$v_i^* = \left(v_i', \text{Vec}(x_i v_i')', \text{Vec}(w_i v_i')', \text{Vec}(q_i v_i')', e_i'\right)'.$$

Since I want to remove (part of) the correlation between treatment and unobserved heterogeneity by including $z_i' v_i$ in the set of regressors, (2.11) suggests to include also v_i^* (instead of including only v_i).

The term $z_i' h_{u;i}$ in (2.8) can now be decomposed as the sum of $v_i^{*\prime}$, $z_i' v_i$ and a remainder which is uncorrelated with both terms

$$\begin{aligned} z_i' h_{u;i} &= v_i'\theta_{u1} + \text{Vec}(x_i v_i')'\theta_{u2} + \text{Vec}(w_i v_i')'\theta_{u3} \\ &\quad + \text{Vec}(q_i v_i')'\theta_{u4} + e_i'\theta_{u5} + z_i' v_i \theta_{u6} + \mathcal{R}_{zu;i}. \end{aligned} \tag{2.12}$$

Unfortunately, we can not rely on Assumption A to assure uncorrelatedness between $(x_i', z_i^{*\prime})$ and $\mathcal{R}_{zu;i}$, but have to assume it explicitly.

Assumption B. The components of $z_i' u_i$ that remain present in the error term are uncor-

[4] This estimator uses $(x_i', z_i^{*\prime}, v_i', z_i' v_i)$ as regressors.

[5] See Appendix A for a formal proof.

related with $(x_i', z_i^{*\prime})$

$$\text{Cov}\left[\mathcal{R}_{zu;i}\left(x_i', z_i^{*\prime}\right)\right] = 0.$$

For a discussion of Assumption B, I refer to subsection 2.2.2.

Taking into account the preceding discussion, the outcome equation (2.8) becomes equal to

$$y_i = s_i'\alpha + \zeta_i, \tag{2.13}$$

where

$$\begin{aligned}
\alpha &= (\tilde{\alpha}_0, \alpha_1', \gamma', \theta_1', \theta_2', \theta_3', \theta_4', \theta_5', \theta_6')' \\
\zeta_i &= \mathcal{R}_{\varepsilon;i} + \mathcal{R}_{zu;i} - \text{E}\left[\mathcal{R}_{zu;i}\right] + z_i'\tilde{u}_i + \tilde{\varepsilon}_i \\
s_i &= \left(1, x_i', z_i^{*\prime}, v_i', \text{Vec}\left(x_i v_i'\right)', \text{Vec}\left(w_i v_i'\right)', \text{Vec}\left(q_i v_i'\right)', e_i', z_i' v_i\right)',
\end{aligned}$$

with $\tilde{\alpha}_0 = \alpha_0 + \text{E}\left[\mathcal{R}_{zu;i}\right]$, $\theta_j = \theta_{\varepsilon j} + \theta_{uj}$, for $j = 1, \ldots, 5$ and $\theta_6 = \theta_{u6}$. Above it was demonstrated that, under Assumption A-B, it holds that $\text{E}\left[\zeta_i s_i\right] = 0$.

Remark 1. The constant term in (2.6), i.e. the parameter α_0, is not identified, unless $\text{E}\left[\mathcal{R}_{zu;i}\right]$ is known.

Since v_i is unobserved and needs to be estimated, the following assumption is made concerning the choice equation (2.2).

Assumption C. $\beta = \text{Vec}\left(\beta_0, B_1', B_2'\right)$ is consistently estimated by some estimate $\hat{\beta}$, for which it holds that

$$\sqrt{N}\left(\hat{\beta} - \beta\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_i + o_p\left(1\right),$$

with $\text{E}\left[\delta_i\right] = 0$,$\text{Var}\left[\delta_i\right] = \Omega$ and Ω a finite and positive definite matrix.

Making use of Assumption C, I define $\hat{s}_i$ as s_i, with v_i replaced by $\hat{v}_i$, and propose to estimate α by

$$\hat{\alpha}_{SE} = \left(\sum_{i=1}^{N}\hat{s}_i\hat{s}_i'\right)^{-1}\sum_{i=1}^{N}\hat{s}_i y_i = \alpha + \left(\sum_{i=1}^{N}\hat{s}_i\hat{s}_i'\right)^{-1}\sum_{i=1}^{N}\hat{s}_i\tilde{\zeta}_i, \tag{2.14}$$

with $\tilde{\zeta}_i$ given by

$$\begin{aligned}\tilde{\zeta}_i &= \zeta_i + (v_i - \hat{v}_i)'\theta_1 + \left(\text{Vec}\left(x_i\,(v_i - \hat{v}_i)'\right)\right)'\theta_2 \\ &\quad + \left(\text{Vec}\left(w_i\,(v_i - \hat{v}_i)'\right)\right)'\theta_3 + \left(\text{Vec}\left(q_i\,(v_i - \hat{v}_i)'\right)\right)'\theta_3 \\ &\quad + \text{Vech}\left(v_i v_i' - \hat{v}_i \hat{v}_i'\right)'\theta_5 + z_i'\,(v_i - \hat{v}_i)\,\theta_6. \end{aligned} \tag{2.15}$$

$\hat{\alpha}_{SE}$ consistently estimates α, by Assumption A-C. Taking into account Lemma A.1 from Appendix A, $\sqrt{N}\,(\hat{\alpha}_{SE} - \alpha)$ is asymptotically normal under standard moment conditions, with variance given by

$$\text{E}\left[s_i s_i'\right]^{-1}\left(\text{E}\left[s_i \zeta_i^2 s_i'\right] + D\Omega D'\right)\text{E}\left[s_i s_i'\right]^{-1}, \tag{2.16}$$

where

$$\begin{aligned} D &= \text{E}\left[s_i\left(\frac{\partial \tilde{\zeta}_i}{\partial \beta'}\right)\right] \\ &= \text{E}\left[s_i\left(\theta_1'\left\{I_L \otimes (1, x_i', w_i')\right\} + \theta_2'\left\{I_L \otimes (x_i\,(1, x_i', w_i'))\right\}\right)\right] \\ &\quad + \text{E}\left[s_i\left(\theta_3'\left\{I_L \otimes (w_i\,(1, x_i', w_i'))\right\} + \theta_4'\left\{I_L \otimes (q_i\,(1, x_i', w_i'))\right\}\right)\right] \\ &\quad + \text{E}\left[s_i\left(-\theta_5'\left\{\Phi_i \otimes (1, x_i', w_i')\right\} + \theta_6'\,(z_i' \otimes (1, x_i', w_i'))\right)\right] \end{aligned} \tag{2.17}$$

with $\Phi_i = (\varphi_{i;1}, \ldots, \varphi_{i;L})$, $\varphi_{i;j} = \text{Vech}\left(\mathbf{1}_{j;L} v_i' + v_i \mathbf{1}_{j;L}'\right)$ and $\mathbf{1}_{j;L}$ the j-th $1 \times L$ unit vector. This variance can be estimated consistently by replacing (α', β') with $\left(\hat{\alpha}_{SE}', \hat{\beta}'\right)$ and expected values with their corresponding population averages. The term $D\Omega D'$ represents the correction to take into account the generated nature of $\hat{v}_i$ (Pierce (1982), Randles (1982), Murphy and Topel (1985)).

2.2.2 Assumption B: Strength and Applicability

Our set of assumptions, the general specification (2.7) combined with Assumption B, is clearly weaker than the linearity assumptions (2.4)-(2.5)[6], since the latter imply that $\mathcal{R}_{zu;i} \equiv 0$, and consequently Assumption B holds.

Assumption B is neither weaker nor stronger than the assumption of constancy of conditional covariances, which reads $\text{E}\left[z_i' u_i \mid x_i, w_i, q_i\right] = \text{E}\left[z_i' u_i\right] = \sigma_{zu}$ (Wooldridge (2003)),

[6] Our set of assumptions is thus a fortiory weaker than Garen's joint normality assumption.

since neither assumption implies the other. To see this, consider

$$\begin{aligned} \mathrm{E}\left[z_i' u_i \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[\mathrm{E}\left[z_i' u_i \mid x_i, w_i, q_i, v_i\right] \mid x_i, w_i, q_i\right] \\ &= \mathrm{E}\left[z_i' h_{u;i} \mid x_i, w_i, q_i\right] \\ &= \mathrm{E}\left[e_i' \theta_{u5} + z_i' v_i \theta_{u6} + \mathcal{R}_{zu;i} \mid x_i, w_i, q_i\right], \end{aligned}$$

where the last equality follows from Assumption A. It is perfectly possible for the right-hand side to be a constant, while each of its terms is a function of (x_i', w_i', q_i'), such that Assumption B is violated. On the other hand, the case where z_i is heteroskedastic, disproves that $\mathrm{E}\left[\mathcal{R}_{zu;i} \mid x_i, w_i, q_i\right] = \mathrm{E}\left[\mathcal{R}_{zu;i}\right]$ implies $\mathrm{E}\left[z_i' u_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[z_i' u_i\right]$. A fortiori, we have that

$$\mathrm{Cov}\left[\mathcal{R}_{zu;i}\left(x_i', z_i^{*\prime}\right)\right] = 0 \quad \not\Rightarrow \quad \mathrm{E}\left[z_i' u_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[z_i' u_i\right].$$

Let us consider now some situations in which our estimator is consistent. A first class of models, for wich our estimator is consistent, has a correlation between z_i and u_i that is a function of (x_i', w_i', q_i'), but this functional dependency is only caused by heteroskedasticity in v_i, with higher order conditional moments uncorrelated with s_i. For this class of models, Assumption B is fulfilled, but Wooldridge's assumption of constancy of conditional covariances is not. As a result, the simple estimator will be consistent, while IV estimation will break down.

A second class of models for which our estimator will produce consistent estimates, is defined by the function $h_u\left(x_i, w_i, q_i, v_i\right) = \zeta v_i$, which is linear in v_i, while $h_\varepsilon\left(x_i, w_i, q_i, v_i\right)$ is still unrestricted. In this case we have that $\mathcal{R}_{zu;i} \equiv 0$, and consequently Assumption B holds. For this class of models the constant term α_0 is identified. A special case of this class of models has (u_i, v_i) normally distributed, with the distribution of ε_i unspecified.

2.3 Comparison with IV and SB

The estimator (2.14) offers some advantages compared to previous estimators. It expands Garen's (1984) SB method to multiple treatments and non-linear relationships between the disturbances. It also improves upon the two-step plug-in estimator of Heckman and Vytlacil (1998) and the IV estimator of Wooldridge (2003), since the assumed dependency between the treatments z_i and their effects $\alpha_{2,i}$ can be tested explicitly with the usual t-test of the parameter θ_4, while not relying on stronger assumptions than the latter two estimators.

A disadvantage common to the proposed estimator, the two-step plug-in estimator and the IV estimator, is that the constant term is not (always) identified. A disadvantage of (2.14), compared to Wooldridge's (2003) IV-estimator, is the prerequisite that the choice equation is linear (or known up to a finite number of parameters), whereas the generation of IVs only requires the first stage regression to estimate linear projections (Wooldridge (2003))[7]. However, when the choice equation is unknown, the constancy of conditional covariances assumption, $\mathrm{E}\left[u_i z_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[u_i z_i\right]$ becomes hard to interpret. In case of a linear projection, the 'remainder'

$$\rho_i \quad = \quad z_i - \mathrm{E}\left[z_i \left(1, x_i', w_i'\right)\right] \left(\mathrm{E}\left[\begin{pmatrix} 1 \\ x_i \\ w_i \end{pmatrix} \left(1, x_i', w_i'\right) \right] \right)^{-1} \begin{pmatrix} 1 \\ x_i \\ w_i \end{pmatrix}$$

is bound to be heteroskedastic (except when the choice equation is linear). Under these circumstances, the constancy of conditional covariances assumption does hold when we make an arcane assumption[8] like $\mathrm{E}\left[\mathrm{E}\left[u_i \mid x_i, w_i, q_i, \rho_i\right] \rho_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[\mathrm{E}\left[u_i \mid x_i, w_i, q_i, \rho_i\right] \rho_i\right]$.

The proposed estimator $\hat{\alpha}_{SE}$ also has a smaller variance than Wooldridge's (2003) IV-estimator. Under constancy of covariances, it holds that

$$\begin{pmatrix} x_i \\ z_i^* \end{pmatrix} - \mathrm{E}\left[\begin{pmatrix} x_i \\ z_i^* \end{pmatrix} \left(1, v_i^{*\prime}\right) \right] \left(\mathrm{E}\left[\begin{pmatrix} 1 \\ v_i^* \end{pmatrix} \left(1, v_i^{*\prime}\right) \right] \right)^{-1} \begin{pmatrix} 1 \\ v_i^* \end{pmatrix} \quad = \quad \begin{pmatrix} x_i - \mathrm{E}\left[x_i\right] \\ z_{i;E}^* - \mathrm{E}\left[z_{i;E}^*\right] \end{pmatrix},$$

which will be denoted $s_i^{(I)}$, with $z_{i;E}^* = \mathrm{Vec}\left(\left(\beta_0 + B_1 x_i + B_2 w_i\right)\left(x_i', w_i', q_i'\right)\right)$, and

$$\tilde{\zeta}_i - \mathrm{E}\left[\tilde{\zeta}_i \left(1, v_i^{*\prime}\right)\right] \left(\mathrm{E}\left[\begin{pmatrix} 1 \\ v_i^* \end{pmatrix} \left(1, v_i^{*\prime}\right) \right] \right)^{-1} \begin{pmatrix} 1 \\ v_i^* \end{pmatrix} \quad = \quad \zeta_i.$$

The proposed estimator of $\left(\alpha_1', \gamma^{*\prime}\right)$ can be written, by the Frisch-Waugh (1933)-Lovell (1963) Theorem, as

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\gamma} - \gamma \end{pmatrix}_R \quad = \quad \left(\sum_{i=1}^{N} s_i^{(I)} s_i^{(I)\prime} \right)^{-1} \sum_{i=1}^{N} s_i^{(I)} \left(\mathcal{R}_{\varepsilon;i} + \mathcal{R}_{zu;i} - \mathrm{E}\left[\mathcal{R}_{zu;i}\right] + z_i' \tilde{u}_i + \tilde{\varepsilon}_i \right).$$

[7] This problem might be solved using nonparametric estimation of the treatment equation (2.2), but not without the introduction of further complications.

[8] This assumption is, for example, fulfilled when $\mathrm{E}\left[u_i \mid x_i, w_i, q_i, \rho_i\right] = f\left(\rho_i\right)$, for which $\mathrm{E}\left[f\left(\rho_i\right) \rho_i \mid x_i, w_i, q_i\right] = 0$.

Comparing this with the IV estimator of (2.6), rewritten in differences from overall means, using $\left(1, x_i', w_i', \text{Vech}\left((x_i', w_i')'(x_i', w_i')\right)', \text{Vec}\left(q_i\,(x_i', w_i')\right)'\right)$ as instruments

$$\begin{pmatrix} \hat{\alpha}_1 - \alpha_1 \\ \hat{\gamma} - \gamma \end{pmatrix}_{IV} = \left(\sum_{i=1}^{N} s_i^{(I)} s_i^{(I)\prime}\right)^{-1} \sum_{i=1}^{N} s_i^{(I)} \left(h_{\varepsilon;i} + z_i' h_{u;i} - \sigma_{uv} + z_i' \tilde{u}_i + \tilde{\varepsilon}_i\right),$$

we see that $\text{Var}\left[(\alpha_1', \gamma')_{SE}\right] \leq \text{Var}\left[(\alpha_1', \gamma')_{IV}\right]$.

2.4 Extension

The derivation in subsection 2.2.1 suggests a way of enhancing our estimator. Suppose $h_u\,(x_i, w_i, q_i, v_i)$ is restricted to be a polynomial in x_i, w_i, q_i and v_i of maximum degree K

$$h_u\,(x_i, w_i, q_i, v_i) = \sum_{k=0}^{K} \sum_{|\lambda|=k} c_\lambda\,(x_i', w_i', q_i', v_i')^\lambda + \mathcal{R}_{zu;i},$$

with $\mathcal{R}_{zu;i} \equiv 0$, where the multi-index λ denotes $\lambda = (\lambda_1, \ldots, \lambda_{K_1+K_2+K_3+L})$ and where we defined $|\lambda| = \lambda_1 + \ldots + \lambda_{K_1+K_2+K_3+L}$. When terms of degree higher than one in v_i are orthogonalized with respect to all monomials[9] of degree lower than or equal to K in the exogenous variables x_i, w_i and q_i, then terms of degree zero in v_i will be lacking h_u, by Assumption A.

Conditional expectation $h_\varepsilon\,(x_i, w_i, q_i, v_i)$ is unrestricted, but assumed to be expandable[10] as a series of monomials in x_i, w_i, q_i and v_i, which can be written as

$$h_\varepsilon\,(x_i, w_i, q_i, v_i) = \sum_{k=0}^{K+1} \sum_{|\lambda|=k} d_\lambda\,(x_i', w_i', q_i', v_i')^\lambda + \mathcal{R}_{\varepsilon;i},$$

with $\mathcal{R}_{\varepsilon;i}$ orthogonal to any polynomial in (x_i', w_i', q_i', v_i') of degree $K+1$ or smaller. Using

[9] As in expression (2.10), where $K = 2$.

[10] By Assumption A, terms of degree zero in v_i will be lacking in this expansion, provided that terms of degree higher than one in v_i are orthogonalized with respect to all monomials of degree lower than or equal to K in the exogenous variables.

the above expressions for h_ε and h_u allows to rewrite outcome equation (2.6) as

$$\begin{aligned} y_i &= \alpha_0 + x_i'\alpha_1 + z_i^{*\prime}\gamma + \sum_{k=0}^{K}\sum_{|\lambda|=k} z_i' c_\lambda \left(x_i', w_i', q_i', v_i'\right)^\lambda \\ &\quad + \sum_{k=0}^{K+1}\sum_{|\lambda|=k} d_\lambda \left(x_i', w_i', q_i', v_i'\right)^\lambda + \mathcal{R}_{\varepsilon;i} + z_i'\tilde{u}_i + \tilde{\varepsilon}_i. \end{aligned} \tag{2.18}$$

By expanding the original set of regressors $(1, x_i', z_i^{*\prime})$, with $z_i h_u\left(x_i, w_i, q_i, v_i\right)$ and with all terms up to degree $(K+1)$ of the series expansion of $h_\varepsilon\left(x_i, w_i, q_i, v_i\right)$, Assumption B will automatically be fulfilled, since $\mathcal{R}_{zu;i} \equiv 0$. Consequently estimation of (2.18) by OLS will be consistent.

2.5 Simulations and Practical Issues

In this section, I present the results of a small Monte Carlo simulation[11]. Each Data Generating Process (DGP) consists of the outcome equation (2.1), the choice equation (2.2), with scalar choice variable z_i, i.e. $L = 1$, and of the treatment equation (2.3), with $\Gamma_3 = 0$

$$\begin{aligned} y_i &= \alpha_0 + x_i'\alpha_1 + z_i'\alpha_{2,i} + \varepsilon_i \\ z_i &= \beta_0 + \beta_1 x_i + \beta_2 w_i + v_i, \\ \alpha_{2,i} &= \gamma_0 + \gamma_1 x_i + \gamma_2 w_i + u_i. \end{aligned}$$

In each DGP, the vector $(x_i, w_i, \psi_i, \tilde{u}_i, \tilde{\varepsilon}_i)$ is generated as $NID\left(0, I_5\right)$. The number of individuals was taken to be 1000, and the number of replications was 10000. The IV estimator was obtained by estimating (2.6), using $\left(1, x_i', w_i', \mathrm{Vech}\left(x_i x_i'\right)', \mathrm{Vec}\left(x_i w_i'\right)'\right)$ as instruments. The SB estimator was obtained as described in section (2.1) and $\hat{\alpha}_{SE}$ is given by (2.14).

The DGPs considered differ in the way the errors ε_i, u_i and v_i are related.

DGP 1: $v_i = \psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + \tilde{\varepsilon}_i$,

DGP 2: $v_i = \psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + v_i^2 - \mathrm{E}\left[v_i + v_i^2 \mid x_i, w_i, q_i\right] + \tilde{\varepsilon}_i$,

DGP 3: $v_i = \psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + v_i^2 + v_i^3 - \mathrm{E}\left[v_i + v_i^2 + v_i^3 \mid x_i, w_i, q_i\right] + \tilde{\varepsilon}_i$,

[11] This simulation was carried out in *R* (*R* Development Core Team (2004)).

	DGP 1			*DGP* 2			*DGP* 3	
	IV	*SB*	*SE*	*IV*	*SB*	*SE*	*IV*	*SE*
1	1.997 (0.143)	0.999 (0.101)	1.000 (0.123)	1.999 (0.141)	0.555 (0.119)	0.011 (0.166)	2.000 (0.140)	0.016 (0.349)
x	0.998 (0.149)	1.000 (0.099)	1.000 (0.109)	0.998 (0.148)	1.001 (0.114)	1.001 (0.134)	0.999 (0.148)	0.999 (0.187)
z	1.003 (0.113)	1.001 (0.081)	1.000 (0.085)	1.002 (0.113)	1.111 (0.092)	0.998 (0.100)	1.001 (0.112)	1.000 (0.127)
zx	0.999 (0.089)	1.000 (0.068)	1.000 (0.073)	0.998 (0.088)	0.889 (0.081)	1.000 (0.078)	0.997 (0.087)	1.001 (0.110)
zw	0.997 (0.088)	0.999 (0.067)	0.999 (0.071)	0.998 (0.087)	0.889 (0.080)	1.001 (0.077)	0.997 (0.087)	1.000 (0.109)
v	/	0.997 (0.099)	0.998 (0.116)	/	0.332 (0.121)	0.008 (0.154)	/	2.969 (0.393)
zv	/	1.000 (0.058)	1.000 (0.083)	/	1.555 (0.066)	1.992 (0.120)	/	1.985 (0.317)

Table 2.1: Estimation results: homoskedastic v_i, different specifications for $h_{\varepsilon;i}$, $h_{u;i} = v_i$.

DGP 4: $v_i = x_i\psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + \tilde{\varepsilon}_i$,

DGP 5: $v_i = w_i\psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + \tilde{\varepsilon}_i$,

DGP 6: $v_i = x_i\psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + v_i^2 - \mathrm{E}\left[v_i + v_i^2 \mid x_i, w_i\right] + \tilde{\varepsilon}_i$,

DGP 7: $v_i = w_i\psi_i$, $u_i = v_i + \tilde{u}_i$ and $\varepsilon_i = v_i + v_i^2 - \mathrm{E}\left[v_i + v_i^2 \mid x_i, w_i\right] + \tilde{\varepsilon}_i$.

In Tables 2.1-2.3 information about the first two empirical moments of the 10000 replications for each estimator are given in the format mean (standard deviation). For $\hat{\alpha}_{SE}$ only a limited number of parameters are displayed.

In Table 2.1 the results of the first three DGPs are presented. DGP 1 is the benchmark case: all three estimators are consistent. The claim that $\mathrm{Var}\left[(\alpha_1', \gamma')_{SE}\right] \leq \mathrm{Var}\left[(\alpha_1', \gamma')_{IV}\right]$ from Section 2.3 is substantiated by the result for DGP 1. In DGP 2 and 3, where $h_{\varepsilon;i}$ is no longer linear, but quadratic and cubic, IV and SE remain consistent, but SB breaks down.

Table 2.2 displays the result of the two DGPs with heteroskedastic errors v_i. In these cases IV is clearly inconsistent, while both the SB estimator and $\hat{\alpha}_{SE}$ perform excellent, with standard errors which are only slightly larger than in the benchmark case.

The results for a nonlinear $h_{\varepsilon;i}$ combined with heteroskedastic errors v_i, DGP 6 and 7, are given in Table 2.3. The only estimator that remains consistent is $\hat{\alpha}_{SE}$. Some further remarks on these results are in order.

Preliminary Monte Carlo simulations indicated a performance of our estimator (2.14) as was to be expected on theoretical grounds in case of a heteroskedastic z_i. However, when the DGP modelled z_i as homoskedastic, severe ill-conditioning or multicollinearity arose in the data matrix ,$(\hat{s}_1, \ldots, \hat{s}_N)$, of the outcome equation (2.1). The latter condition

	DGP 4			*DGP* 5		
	IV	*SB*	*SE*	*IV*	*SB*	*SE*
1	1.257 (0.203)	0.998 (0.097)	1.002 (0.108)	2.236 (0.272)	0.997 (0.117)	1.005 (0.190)
x	0.014 (0.347)	1.002 (0.114)	1.002 (0.177)	1.973 (0.317)	0.999 (0.114)	1.004 (0.173)
z	1.247 (0.174)	1.001 (0.088)	0.997 (0.093)	0.268 (0.260)	1.002 (0.094)	0.997 (0.141)
zx	1.735 (0.190)	1.001 (0.085)	0.998 (0.124)	0.758 (0.127)	1.001 (0.074)	1.000 (0.077)
zw	0.753 (0.125)	1.000 (0.075)	0.999 (0.074)	1.730 (0.196)	1.001 (0.084)	1.004 (0.133)
v	/	0.995 (0.134)	1.001 (0.161)	/	0.993 (0.118)	1.008 (0.256)
zv	/	1.001 (0.061)	1.002 (0.131)	/	1.000 (0.061)	0.995 (0.143)

Table 2.2: Estimation results: heteroskedastic v_i, $h_{\varepsilon;i} = v_i$, $h_{u;i} = v_i$.

	DGP 6			*DGP* 7		
	IV	*SB*	*SE*	*IV*	*SB*	*SE*
1	1.257 (0.206)	0.866 (0.139)	1.008 (0.117)	2.243 (0.277)	0.270 (0.230)	1.003 (0.197)
x	0.016 (0.348)	1.585 (0.230)	1.003 (0.178)	1.980 (0.326)	0.402 (0.248)	0.996 (0.179)
z	1.247 (0.177)	1.030 (0.126)	0.999 (0.091)	0.261 (0.268)	1.626 (0.221)	1.004 (0.145)
zx	1.735 (0.189)	0.386 (0.160)	0.999 (0.138)	0.755 (0.127)	0.971 (0.105)	1.000 (0.079)
zw	0.752 (0.125)	0.973 (0.102)	1.001 (0.078)	1.733 (0.200)	0.383 (0.164)	0.998 (0.137)
v	/	0.220 (0.191)	1.007 (0.178)	/	−0.376 (0.276)	1.003 (0.266)
zv	/	1.746 (0.084)	0.994 (0.153)	/	1.747 (0.085)	0.994 (0.150)

Table 2.3: Estimation results: heteroskedastic v_i, $h_{\varepsilon;i} = v_i + v_i^2 - \mathrm{E}\left[v_i + v_i^2 \mid x_i, w_i, q_i\right]$, $h_{u;i} = v_i$.

can be detected by decomposing the cross-product matrix, C, of the regressors in function of its eigenvalues and eigenvectors as

$$C = \sum_{i=1}^{N} \hat{s}_i \hat{s}_i' = \Xi \Lambda \Xi',$$

with Λ the (diagonal) matrix of eigenvalues and Ξ the matrix of eigenvectors of C, i.e. they are the solution to $C\Xi = \Xi\Lambda$. Eigenvalues λ_j for which the condition index $\eta_j = \sqrt{\max_k [\lambda_k] / \lambda_j}$ is larger than some treshold T, contribute to the multicollinearity (Belsley, Kuh and Welsch (1980), Sen and Srivastava (1990)). The proportion of the variance of the l-th component of $\hat{\alpha}_{SE}$ contributed by the j-th eigenvalue is given by

$$V_{lj} = \frac{\lambda_j^{-1} \gamma_{lj}^2}{\sum_{k=1}^{K_1+K_2+1} \lambda_k^{-1} \gamma_{lk}^2}.$$

Although I am fully aware that the removal of multicollinear variables requires judgment of the researcher and should not be relegated to automated procedures, I implemented the following variable removal method, in order to carry out a Monte Carlo simulation:

- If $\max_j [\eta_j] \geq T$, then compute $k = \arg\max_j [\eta_j]$, i.e. the k-th eigenvalue has the highest condition number.
- Compute $m = \arg\max_{l \in \mathcal{L}} [V_{lk}]$. For each variable l, I compute the fraction of its variance that is 'caused' by eigenvalue k. The variable with the highest contribution from eigenvalue k is subsequently removed. The set $\mathcal{L}$ comprises only the indices of the following subvector $(\hat{v}_i x_i', \hat{v}_i w_i', \hat{v}_i q_i', e_i)$ of $\hat{s}_i$. These are exactly the variables I added extra to the outcome equation, compared with the SB method, which did not seem to be affected by multicollinearity.

I computed the cut-off value T as

$$T = \arg [F(x) = 1 - G(x)],$$

where F is the CDF of $\max_j [\eta_j]$ under DGP1 and G the CDF of $\max_j [\eta_j]$ under DGP 4. This value was approximately equal to 50. It was chosen because under DGP 1 the ill-conditioning of the data matrix was always caused by one sole small condition index,

while a too low treshold caused bias in $\hat{\alpha}_{SE}$ under DGP 4. A treshold value equal to $\arg\left[F\left(x\right)=1-G\left(x\right)\right]$ equalises the probability of 'incorrectly keeping a variable' in the outcome equation under DGP 1, with the probability of 'incorrectly dropping a variable' from the outcome equation under DGP 4. The CDF of $\max_j\left[\eta_j\right]$ seemed rather robust with respect to the specification of the heteroskedasticity. Alternative specifications like $v_i=\psi_i\sqrt{|x_i|}$ resulted in approximately the same value for T.

2.6 Conclusion

In this chapter an estimator for a CRC model is presented that is an extension of Garen's (1984) SB method, but does not assume a linear relationship between the error term of the choice equation and the unobserved heterogeneity. I propose to include more terms in the set of regressors compared to the SB method. The resulting outcome equation is subsequently estimated by OLS. The proposed estimator allows a test of the correlation between treatment and effect and has smaller variance compared to Wooldridge's (2003) IV estimator.

Chapter 3

Semiparametric Estimation of the Correlated Random Coefficient Model

Abstract. In this paper I discuss semiparametric estimation of the outcome equation of the Correlated Random Coefficient (CRC) model (Heckman and Vytlacil (1998)). By making use of the unified treatment of models for truncation, sample selection and limited dependent variables (Heckman (1976)) and by including an unknown function of the generalized residuals (Cox and Snell (1968), Gouriéroux et al. (1987)) in the outcome equation, I propose to estimate the latter by series regression, resulting in $\sqrt{N}$-consistent estimation of the parameters of interest. As such, it can be considered as an extension of Garen's Selectivity Bias Method (1984).
Keywords: Average treatment effect; Correlated random coefficient model; Sample selection; Semiparametric estimation; Unobserved heterogeneity.
JEL classification: C14, C21, C22, C24.

3.1 Introduction

The Correlated Random Coefficient (CRC) model is a model in which some coefficients are stochastically dependent on some continuous choice or treatment variables present in the outcome equation (Heckman and Vytlacil (1998)). In the recent human capital literature it takes a central place (Card (2001)) Several estimation methods have been proposed to estimate the CRC model: two-stage plug-in estimation (Heckman and Vytlacil (1998)), instrumental variables estimation (Wooldridge (2003,1997)) and Garen's (1984) selectivity

bias approach. The latter method assumes joint normality of all error terms and involves the estimation of the choice equation by least squares. It subsequently incorporates both the choice residual and the interaction term between choice residual and choice variable in the outcome equation and estimates this extended outcome equation by OLS. I will extend the latter method by dropping the assumption of joint normality and by semiparametrically estimating the outcome equation.

Although the vast majority of empirical applications in econometrics still relies on simple parametric techniques, nonparametric methods have attracted attention since quite some time. While early work involved mostly density estimation (Rosenblatt (1956)), nonparametric regression became increasingly investigated (Stone (1985), White (1980a)). See Härdle, Liang and Gao (2000) and Yatchew (2003) for recent overviews. Some disadvantages of the fully nonparametric approach, such as the need for large datasets and the curse of dimensionality, can be alleviated by considering additive stochastic regression models (Fan, Härdle and Mammen (1998), Linton (1997), Stone (1985)), but most improvements in this area come from partially linear or semiparametric models.

It is well established that the parameter vector of the parametric part of a semilinear model can be $\sqrt{N}$-consistently estimated, when the data consists of an IID random sample (see Ai and McFadden (1997), Chen (1988), Li (1996), Newey (1997), Robinson (1988)), or when they are heteroskedastically independently distributed (Andrews (1991), Donald and Newey (1994), Fan et al. (1995)). These results have frequently been applied to different models involving sample selection (see Vella (1998) for an overview) and other simultaneous equation models.

Ahn and Powell (1993), Das et al. (2003) and Newey (1999) semiparametrically estimate different models involving sample selection, whereby they all estimate the first stage nonparametrically under an index restriction (Ichimura (1993)). Lee (1994), Chen (1997) and Christofides et al. (2003) consider sample selection models subject to Tobit-type selection rules. While the latter semiparametrically extend a two-stage estimator proposed by Vella (1998, 1992) and Wooldridge (1994), by including an unknown function of a tobit type residual in the outcome equation, Lee (1994) and Chen (1997) develop semiparametric two-stage methods without a priori knowledge of error distributions. Newey et al. (1999) discuss two-step nonparametric estimation of a triangular simultaneous equation model. Das (2005), finally, develops a two-step estimator whereby an unknown function of the nonparametricaly estimated propensity score, appears in the outcome eqation and is estimated by series regression. In this chapter, I will apply previous results on semiparametric

estimation to the CRC model and extend it in several directions.

Making use of the common structure of statistical models of truncation, sample selection and limited dependent variables (Heckman (1979, 1976)), I allow the choice equation to be either linear or nonlinear and the endogenous variable to be either continuous, truncated or qualitative. Taking into account the results of Newey et al. (1990), which suggest that correct specification of the outcome equation is of higher importance than the specification of the error distribution, I assume the choice equation to be known. This approach is comparable with Christofides et al. (2003), Vella (1992) and Wooldridge (1994).

Secondly, no a priori assumptions are made about the functional form of the correlation between the choice error term on the one hand and both the correlated coefficient and the outcome error term on the other hand. To correct for endogeneity and/or selection bias, the outcome equation will be expanded with two unknown functions in the generalized residual of the choice equation (Cox and Snell (1968), Gouriéroux et al. (1987)), one of which interacted with the choice variable. In order to estimate the outcome equation semiparametrically, I adapt the series estimator of Donald and Newey (1994) to estimated regressors.

In section 3.2 a fairly general correlated coefficient model is presented, together with its assumptions. Section 3.3 discusses semiparametric estimation of this CRC model. Section 3.4 concludes and proofs are presented in Appendix B.

3.2 Theoretical Framework

Consider the following model, which consists of the outcome (3.1), the reduced form of the endogenous choice or treatment (3.2), the censoring and selection rules (3.3)-(3.4) and an equation for the heterogeneous treatment effect $\alpha_{2,i}$ in terms of a vector of observed variables (3.5),

$$
\begin{aligned}
y_i^* &= x_i'\alpha_1 + z_i\alpha_{2,i} + \varepsilon_i, && (3.1)\\
z_i^* &= f_1\left(x_i, w_i; \beta\right) + v_i, && (3.2)\\
z_i &= f_2\left(z_i^*\right), && (3.3)\\
y_i &= f_3\left(y_i^*, z_i\right), && (3.4)\\
\alpha_{2,i} &= x_i'\gamma_1 + w_i'\gamma_2 + q_i'\gamma_3 + u_i, && (3.5)
\end{aligned}
$$

where $i = 1,\ldots,N$ and all variables are assumed to be independently distributed across individuals. Making abstraction of the parameter heterogeneity, model (3.1)-(3.3) is a special case of Vella's (1993) model combining endogeneity and sample selection, but here (3.2) is allowed to be nonlinear. Model (3.1)-(3.4) is the cross-section equivalent of the panel data model considered by Vella and Verbeek (1999), except that I restrict (3.1) to be linear.

The scalar latent endogenous variables y_i^* and z_i^* have observed counterparts y_i and z_i. While the former is the observed outcome, we are particularly interested in the mean influence z_i exerts on it. The function $f_2(z_i^*)$ is allowed to be equal to one of the following

- $f_2(z_i^*) = z_i^*$,
- $f_2(z_i^*) = \mathbf{1}_{z_i^*>0}$ or
- $f_2(z_i^*) = z_i^* \mathbf{1}_{z_i^*>0}$,

in case z_i is a continuous, a dummy or a censored endogenous regressor[1]. Likewise, the allowed possibilities for (3.4) are

- $f_3(y_i^*, z_i) = y_i^* \mathbf{1}_{z_i>0}$, with $y_i = 0$ unobserved, or
- $f_3(y_i^*, z_i) = y_i^*$,

whether y_i is incidentally truncated or not. The rather broad specification (3.1)-(3.5) includes the linear CRC model of Heckman and Vytlacil[2] (1998), but also allows for a qualitative or censored endogenous variable z_i (Vella (1993)) and for selection bias in the primary equation (Heckman (1979,1976), Vella and Verbeek (1999)).

The observed vector x_i, which possibly contains the constant term, has dimension K_1, and influences the outcome linearly with unknown parameter vector α_1. The scalar choice variable z_i is assumed to behave according to the reduced form (3.2), where the vector w_i has dimension $K_2 > 0$. This strict inequality is the usual identification restriction, which allows the use of the vector w_i as a set of instruments for the treatment z_i in case f_1 is linear. The variables in (x_i', w_i') are strictly exogenous in a sense to be described later (see Assumption A below). The heterogeneous treatment effects $\alpha_{2,i}$ are allowed to consist of a linear combination of (x_i', w_i', q_i'), a vector of observed variables and an error term, u_i

[1] $\mathbf{1}_A$ denotes the indicator function of the event A (Christofides et al. (2003)).

[2] In the linear CRC model of Heckman and Vytlacil (1998) we have that $y_i = y_i^*$, $z_i = z_i^*$ and $f_1 = \beta_0 + x_i'\beta_1 + w_i'\beta_2$.

(Garen (1984), Heckman and Vytlacil (1998)) as in (3.5), where the vector q_i has dimension $K_3 \geq 0$. The parameter of interest is the average treatment effect (ATE),

$$\begin{aligned} \alpha_2 &\equiv \mathrm{E}\left[\alpha_{2,i}\right] \\ &= \mathrm{E}\left[x_i'\right]\gamma_1 + \mathrm{E}\left[w_i'\right]\gamma_2 + \mathrm{E}\left[q_i'\right]\gamma_3, \end{aligned}$$

which suggests to rewrite the treatment effects $\alpha_{2,i}$ as $\gamma_0^* + \tilde{x}_i'\gamma_1 + \tilde{w}_i'\gamma_2 + \tilde{q}_i'\gamma_3 + u_i$, with $\gamma_0^* = \mathrm{E}\left[\alpha_{2,i}\right]$, $\tilde{x}_i = x_i - \mathrm{E}\left[x_i\right]$, and similarly for the other variables, since then the parameter of interest, α_2, and its standard errors can immediately be retrieved.

This version of the CRC model is completed with the following assumption about the variables present in (3.1)-(3.5).

Assumption A. The vector of observations $(x_i', w_i', q_i', y_i, z_i)$ is independently distributed across i, and the variables in (x_i', w_i', q_i') are strictly exogenous in the model (3.1)-(3.5), i.e.

$$\mathrm{E}\left[\varepsilon_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[u_i \mid x_i, w_i, q_i\right] = \mathrm{E}\left[v_i \mid x_i, w_i, q_i\right] = 0.$$

Under the above Assumption A, the choice variable z_i is allowed to be endogenous and possibly qualitative or truncated, and its individual-specific return $\alpha_{2,i}$ is allowed to have an unobserved component, which is correlated with unobserved components in the choice variable. For a theoretical justification of the CRC model in the context of the schooling choice and its returns see section 1.2 or Card (2001). The aim of this chapter is to examine how the parameter of interest, α_2, can be consistently estimated, while making the least possible assumptions about the functional form of the dependency between ε_i, u_i and v_i.

3.3 Estimation

The latent outcome y_i^* from (3.1) can be rewritten in terms of observed variables as

$$y_i^* = x_i'\alpha_1 + z_i^{*\prime}\gamma + z_i u_i + \varepsilon_i, \tag{3.6}$$

where $z_i^* = z_i\left(x_i', w_i', q_i'\right)'$ and $\gamma = \left(\gamma_1', \gamma_2', \gamma_3'\right)'$. In the absence of endogeneity and sample selection, the parameter vectors α_1 and γ could be $\sqrt{N}$-consistently estimated by least squares under suitable conditions. To cope with the problems of endogeneity and

sample selection, consider now the generalized residuals of the choice equation[3], $\tilde{v}_i = \tilde{v}_i(x_i, w_i, z_i, \beta) = \mathrm{E}[v_i | x_i, w_i, z_i]$, in the sense of Gouriéroux et al. (1987)[4], and assume the following.

Assumption B. The conditional expectations

$$\begin{aligned} \mathrm{E}[\varepsilon_i | x_i, w_i, q_i, z_i] &= \mathrm{E}[\varepsilon_i | \tilde{v}_i] = g_\varepsilon(\tilde{v}_i), \\ \mathrm{E}[u_i | x_i, w_i, q_i, z_i] &= \mathrm{E}[u_i | \tilde{v}_i] = g_u(\tilde{v}_i), \end{aligned}$$

exist.

While the general specification of our model (3.1)-(3.5) generalizes Garen's (1984) SB method to allow for a censored or qualitative choice variable, Assumption B generalizes Garen's (1984) SB method to nonlinear relationships between the choice error and the outcome error, respectively the treatment effect error. It requires that, conditional on $\tilde{v}_i$, ε_i and u_i are mean independent of q_i and are uncorrelated with all functions of (x_i', w_i', z_i) other than $g_{\varepsilon;i}$, respectively $g_{u;i}$. It also allows for a heteroskedastic specification of z_i. When y_i is censored, Assumption B is equivalent to the index restriction (Vella (1998)), used in semiparametric estimation of sample selection models, since $\tilde{v}_i$ is an invertible function of the single index $f_1(x_i, w_i; \beta)$. When z_i is categorical and y_i is uncensored, however, Assumption B is not equivalent to the index restriction.

In contrast with the literature on semiparametric estimation of sample selection models, I do assume that the marginal distribution of v_i is known, up to the parameter vector β. I do not make any supplementary assumptions on the joint distribution of $(\varepsilon_i, u_i, v_i)$ other than Assumptions A-B, however.

When $(\varepsilon_i, u_i, v_i)$ is Gaussian, we have that $g_\varepsilon(\tilde{v}_i) = \sigma_{\varepsilon v}\tilde{v}_i$ and $g_u(\tilde{v}_i) = \sigma_{uv}\tilde{v}_i$. If the endogenous variable is continuous and uncensored, i.e. $z_i = z_i^*$, the generalized residuals are equal to the residuals in the usual sense, i.e. $\tilde{v}_i = v_i$. If z_i is a binary choice variable, the generalized residuals are equal to the inverse of Mills ratio (Heckman (1979,1976)).

A drawback of Assumption B is that the conditional mean of the errors ε_i and v_i only depend on the generalized residual $\tilde{v}_i$. In general we would have that $\mathrm{E}[\varepsilon_i | x_i, w_i, q_i, z_i] = h_\varepsilon(x_i, w_i, q_i, z_i)$ and $\mathrm{E}[u_i | x_i, w_i, q_i, z_i] = h_u(x_i, w_i, q_i, z_i)$, with $h_\varepsilon(\cdot)$ and $h_u(\cdot)$ only restricted by Assumption A. Assumption B, however, restricts the conditional expectations

[3] For the logit model, for instance, the generalized residual is given by $\tilde{v}_i = z_i - (1 + \exp(-p_i'\beta))^{-1}$, for the probit model by $\tilde{v}_i = \frac{\varphi(p_i'\beta)[z_i - \Phi(p_i'\beta)]}{\Phi(p_i'\beta)[1 - \Phi(p_i'\beta)]}$ and for the linear model by $\tilde{v}_i = v_i = z_i - p_i'\beta$.

[4] See also Cox and Snell (1968), who introduced the concept of generalized residual.

of ε_i and v_i to depend neither on q_i, nor on functions of (x_i, w_i, z_i) other than $\tilde{v}_i$. This restriction would also be implied by independence between regressors and errors (Newey (1999), Powell (1994)).

The main advantage of Assumption B lies in the fact that it allows to simultaneously treat censored and uncensored outcomes, as well as continuous, discrete and limited dependent endogenous variables.

Using Assumption B , the latent outcome variable y_i^* from (3.6) can be reformulated as

$$y_i^* = x_i'\alpha_1 + z_i^{*\prime}\gamma + z_i g_u(\tilde{v}_i) + g_\varepsilon(\tilde{v}_i) + \xi_i, \tag{3.7}$$

where $\mathrm{E}[\xi_i | x_i, w_i, q_i, z_i] = 0$. A sufficient condition for identification of (3.7) is considered in the following proposition (Röhrig (1988)), which is analogous to Theorem 2.3 from Das et al. (2003).

Proposition 3.1. *Suppose $g_\varepsilon(\tilde{v}_i)$, $g_u(\tilde{v}_i)$ are continuously differentiable and $\tilde{v}_i(x_i, w_i, z_i)$ is continuously differentiable with respect to x_i and w_i, all with continuous distribution functions almost everywhere. If z_i is not constant over the sample and rank $\left[\partial(z_i, \tilde{v}_i)'/\partial w_i\right] = 2$, or if z_i is constant over the sample, rank $[\partial\tilde{v}_i/\partial w_i] = 1$ and $\gamma_1 = 0$, then $(\alpha_1, \gamma_1, \gamma_2, \gamma_3)$ is identified.*

Proof. Suppose first that z_i is not constant over the sample. For any observationally equivalent model $(\check{\alpha}_1, \check{\gamma}_1, \check{\gamma}_2, \check{\gamma}_3, \check{g}_u(\tilde{v}_i), \check{g}_\varepsilon(\tilde{v}_i))$, it must hold that

$$\mathrm{E}[y_i^* | x_i, w_i, q_i, z_i] = x_i'\check{\alpha}_1 + (z_i x_i')\check{\gamma}_1 + (z_i w_i')\check{\gamma}_2 + (z_i q_i')\check{\gamma}_3 + z_i\check{g}_u(\tilde{v}_i) + \check{g}_\varepsilon(\tilde{v}_i),$$

which, together with (3.7) implies that

$$\begin{aligned} &x_i'(\check{\alpha}_1 - \alpha_1) + (z_i x_i')(\check{\gamma}_1 - \gamma_1) + (z_i w_i')(\check{\gamma}_2 - \gamma_2) \\ &+ (z_i q_i')(\check{\gamma}_3 - \gamma_3) + z_i(\check{g}_u(\tilde{v}_i) - g_u(\tilde{v}_i)) + (\check{g}_\varepsilon(\tilde{v}_i) - g_\varepsilon(\tilde{v}_i)) = 0. \end{aligned} \tag{3.8}$$

Differentiating this last identity with respect to x_i, w_i, q_i and $\tilde{v}_i$ results in

$$
\begin{aligned}
0 &= (\check{\alpha}_1 - \alpha_1) + z_i (\check{\gamma}_1 - \gamma_1) + \left(\frac{\partial z_i}{\partial x_i}, \quad \frac{\partial \tilde{v}_i}{\partial x_i} \right) \\
&\quad \times \begin{pmatrix} x_i' (\check{\gamma}_1 - \gamma_1) + w_i' (\check{\gamma}_2 - \gamma_2) + q_i' (\check{\gamma}_3 - \gamma_3) + (\check{g}_u (\tilde{v}_i) - g_u (\tilde{v}_i)) \\ z_i \left(\frac{\partial \check{g}_u(\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_u(\tilde{v}_i)}{\partial \tilde{v}_i} \right) + \left(\frac{\partial \check{g}_\varepsilon(\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_\varepsilon(\tilde{v}_i)}{\partial \tilde{v}_i} \right) \end{pmatrix}, \qquad (3.9)
\end{aligned}
$$

$$
\begin{aligned}
0 &= z_i (\check{\gamma}_2 - \gamma_2) + \left(\frac{\partial z_i}{\partial w_i}, \quad \frac{\partial \tilde{v}_i}{\partial w_i} \right) \\
&\quad \times \begin{pmatrix} x_i' (\check{\gamma}_1 - \gamma_1) + w_i' (\check{\gamma}_2 - \gamma_2) + q_i' (\check{\gamma}_3 - \gamma_3) + (\check{g}_u (\tilde{v}_i) - g_u (\tilde{v}_i)) \\ z_i \left(\frac{\partial \check{g}_u(\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_u(\tilde{v}_i)}{\partial \tilde{v}_i} \right) + \left(\frac{\partial \check{g}_\varepsilon(\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_\varepsilon(\tilde{v}_i)}{\partial \tilde{v}_i} \right) \end{pmatrix}, \qquad (3.10)
\end{aligned}
$$

$$
0 = z_i (\check{\gamma}_3 - \gamma_3), \qquad (3.11)
$$

$$
0 = z_i \left(\frac{\partial \check{g}_u (\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_u (\tilde{v}_i)}{\partial \tilde{v}_i} \right) + \left(\frac{\partial \check{g}_\varepsilon (\tilde{v}_i)}{\partial \tilde{v}_i} - \frac{\partial g_\varepsilon (\tilde{v}_i)}{\partial \tilde{v}_i} \right) \qquad (3.12)
$$

with probability one. We now have that from (3.11) follows that $\check{\gamma}_3 = \gamma_3$. Furthermore, (3.12) implies that $\check{g}_u (\tilde{v}_i) = g_u (\tilde{v}_i) + C_1$ and $\check{g}_\varepsilon (\tilde{v}_i) = g_\varepsilon (\tilde{v}_i) + C_2$. However, from Assumption A we have that $C_1 = C_2 = 0$. Equation (3.10) now implies $\check{\gamma}_1 = \gamma_1$ and $\check{\gamma}_2 = \gamma_2$. Condition (3.8) finally implies $x_i' (\check{\alpha}_1 - \alpha_1) = 0$, which identifies α.

Suppose now that $z_i = 1$ over the sample. Condition (3.12) implies that $g_u (\tilde{v}_i)$ and $g_\varepsilon (\tilde{v}_i)$ are not identified. Dropping g_u results in the following condition for any observationally equivalent model

$$
\begin{aligned}
x_i' (\check{\alpha}_1 - \alpha_1) + x_i' (\check{\gamma}_1 - \gamma_1) + w_i' (\check{\gamma}_2 - \gamma_2) & \\
+ q_i' (\check{\gamma}_3 - \gamma_3) + (\check{g}_\varepsilon (\tilde{v}_i) - g_\varepsilon (\tilde{v}_i)) &= 0.
\end{aligned}
$$

Again we have that $\check{g}_\varepsilon (\tilde{v}_i) = g_\varepsilon (\tilde{v}_i)$. If now $\check{\gamma}_1 = \gamma_1 = 0$, then $(\alpha_1, \gamma_2, \gamma_3)$ is identified, since q_i and w_i do not contain variables in common with x_i. □

Proposition 3.1 implies that incidental truncation of the outcome based on a binary endogenous variable that is also included in the outcome equation, results in an unidentified model, except in the rare case that the return of this endogenous selection variable is a function of exogenous variables (not including the constant term). However, all other

combinations of censoring rule f_2 (3.3) and selection rule f_3 (3.4), as listed in section 3.2, are allowed. The rank condition of Proposition 3.1 requires that $K_2 \geq 1$, if z_i is constant over the whole sample and that $K_2 \geq 2$ otherwise. Remark that this is a sufficient condition, which might be redundant for identification (see Newey et al. (1999)).

I obtain an estimator of the parameters (α_1', γ') appearing in (3.7) by regressing y_i on x_i, z_i^*, a first series of polynomials in $\tilde{v}_i$ and z_i times a second series in polynomials in $\tilde{v}_i$, assuming these polynomials exist. This series is meant to approximate the unknown functions $g_\varepsilon(\tilde{v}_i)$ and $g_u(\tilde{v}_i)$ and thus to eliminate the effects of the endogeneity and/or sample selection. The series estimator of Donald and Newey (1994) seems suitable in this context, since it does not restrict the distribution of $\tilde{v}_i$. In addition, it allows for heteroskedasticity.

I assume that the unknown functions $g_\varepsilon(\cdot)$ and $g_u(\cdot)$ from Assumption B can be approximated by the finite series expansions $\sum_{j=1}^{\kappa_\varepsilon(N)} \theta_{\varepsilon j}\varphi_j(\cdot)$ and $\sum_{j=1}^{\kappa_u(N)} \theta_{uj}\varphi_j(\cdot)$. The set $\{\varphi_j(w) : j = 1, 2, \ldots\}$ can be equal to $\{1, w, w^2, w^3, \ldots\}$, but it can also denote the family of orthonormal polynomials with respect to the CDF $F(w)$ (see Szegö (1967)). $\kappa_\varepsilon(N), \kappa_u(N) : \mathbb{N}^+ \to \mathbb{N}^+$ are called truncation sequences. They denote the number of terms to be included in the series for a given sample size N. As a consequence the unknown parameter vectors $\Theta_{\varepsilon\kappa_\varepsilon} = \left(\theta_{\varepsilon 1}, \ldots, \theta_{\varepsilon\kappa_\varepsilon(N)}\right)'$ and $\Theta_{u\kappa_u} = \left(\theta_{u1}, \ldots, \theta_{u\kappa_u(N)}\right)'$ increase in length for increasing sample size.

Define now the vectors that will approximate $g_\varepsilon(\tilde{v}_i)$ and $g_u(\tilde{v}_i)$as

$$\vec{\varphi}_\kappa(\tilde{v}_i) = (\varphi_1(\tilde{v}_i), \ldots, \varphi_\kappa(\tilde{v}_i))'.$$

If the approximation is by power series, it holds that $\vec{\varphi}_\kappa(\tilde{v}_i)$ is equal to $\left(1, \tilde{v}_i, \ldots, \tilde{v}_i^{\kappa(N)}\right)$, where the sample size is assumed to be N. We can now stack the vectors that will approximate $g_\varepsilon(\tilde{v}_i)$ in the matrix

$$\Phi_{N\kappa} = (\vec{\varphi}_{\kappa_\varepsilon}(\tilde{v}_1), \ldots, \vec{\varphi}_{\kappa_\varepsilon}(\tilde{v}_N))',$$

and the vectors that will approximate $z_i g_u(\tilde{v}_i)$ as

$$\Phi^*_{N\kappa} = (z_1\vec{\varphi}_{\kappa_u}(\tilde{v}_1), \ldots, z_N\vec{\varphi}_{\kappa_u}(\tilde{v}_N))'.$$

Remark that $\Phi_{N\kappa}$ and $\Phi^*_{N\kappa}$ do not need to have the same number of columns, since truncation sequences $\kappa_\varepsilon(N), \kappa_u(N)$ are allowed to differ. In general, I let uppercase letters

denote the stacked observations of the corresponding vector of variables.

Since $\tilde{v}_i(\beta)$ is unobserved, it will be estimated by $\hat{\tilde{v}}_i = \tilde{v}_i\left(\hat{\beta}\right)$. Consequently, we need to make some assumptions about the particular estimator $\hat{\beta}$ that will be used to estimate β. More specifically, I assume the existence of a $\sqrt{N}$-consistent estimator of the parameter vector, β, of the secondary equation (3.2) with censoring rule (3.3).

Assumption C. β is consistently estimated by some estimate $\hat{\beta}$, for which it holds that

$$\sqrt{N}\left(\hat{\beta}-\beta\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\delta_i + o_p(1),$$

with $\mathrm{E}[\delta_i]=0$,$\mathrm{Var}[\delta_i]=\Sigma_{\beta\beta}$ and $\Sigma_{\beta\beta}$ a finite and positive definite matrix.

Let now hatted variable vectors and matrices be equal to the corresponding straight ones, with variables depending on $\tilde{v}$ replaced with their observed counterparts (i.e. depending on $\hat{\tilde{v}}$)[5]. For example $\hat{\Phi}_{N\kappa} = \left(\vec{\varphi}_{\kappa_\varepsilon}\left(\hat{\tilde{v}}_1\right),\ldots,\vec{\varphi}_{\kappa_\varepsilon}\left(\hat{\tilde{v}}_N\right)\right)'$. The OLS estimator of $\left(\alpha_1',\gamma',\Theta_{\varepsilon\kappa_\varepsilon}',\Theta_{u\kappa_u}'\right)'$ is given by

$$\begin{pmatrix} \hat{\alpha}_1\left(\hat{\beta}\right) \\ \hat{\gamma}\left(\hat{\beta}\right) \\ \hat{\Theta}_{\varepsilon\kappa_\varepsilon}\left(\hat{\beta}\right) \\ \hat{\Theta}_{u\kappa_u}\left(\hat{\beta}\right) \end{pmatrix} = \left(\hat{R}'\hat{R}\right)^{-1}\hat{R}'Y, \tag{3.13}$$

with $R=(r_1',\ldots,r_N')'$ and $r_i=\left(x_i',z_i^{*\prime},\vec{\varphi}_{\kappa_\varepsilon}(\tilde{v}_i)',z_i\vec{\varphi}_{\kappa_u}(\tilde{v}_i)'\right)'$. The series estimators $\hat{g}_\varepsilon$ of g_ε, respectively $\hat{g}_u$ of g_u, are now given by

$$\begin{aligned} \hat{g}_\varepsilon(v) &= \hat{\vec{\varphi}}_{\kappa_\varepsilon}'(v)\,\hat{\Theta}_{\varepsilon\kappa_\varepsilon} \\ \hat{g}_u(v) &= \hat{\vec{\varphi}}_{\kappa_u}'(v)\,\hat{\Theta}_{u\kappa_u}. \end{aligned}$$

By the Frisch-Waugh (1933)-Lovell (1963) Theorem, the estimator of the parametric

[5] If we approximate the unknown functions by power series, then we have that $\hat{\vec{\varphi}}_\kappa\left(\hat{\tilde{v}}_i\right)$ is equal to $\left(1,\hat{\tilde{v}}_i,\hat{\tilde{v}}_i^2,\ldots,\hat{\tilde{v}}_i^{\kappa(N)}\right)$. However, orthogonal polynomials with respect to the empirical CDF $\hat{F}\left(\hat{\tilde{v}}\right)$, slightly differ from those with respect to the true CDF $F(v)$. In Appendix B it is shown that the former converge to the latter. In the remainder of the text I make abstraction of this complication.

component of y_i may be written as

$$\begin{pmatrix} \hat{\alpha}_1\left(\hat{\beta}\right) \\ \hat{\gamma}\left(\hat{\beta}\right) \end{pmatrix} = \left(S'\mathcal{Q}_{\hat{\Psi}}S\right)^{+} S'\mathcal{Q}_{\hat{\Psi}}Y \tag{3.14}$$

with $S = (s_1', \ldots, s_N')$, $s_i = (x_i', z_i^{*\prime})'$ and $\mathcal{Q}_{\hat{\Psi}} = I - \mathcal{P}_{\hat{\Psi}}$, where

$$\mathcal{P}_{\hat{\Psi}} = \hat{\Psi}_{N\kappa}\left(\hat{\Psi}_{N\kappa}'\hat{\Psi}_{N\kappa}\right)^{-1}\hat{\Psi}_{N\kappa}',$$

denotes the projection matrix on the column space of $\hat{\Psi}_{N\kappa} = \begin{pmatrix} \hat{\Phi}_{N\kappa}, & \hat{\Phi}_{N\kappa}^{*} \end{pmatrix}$. The estimator (3.14) can also be rewritten as

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}_1\left(\hat{\beta}\right) - \alpha_1 \\ \hat{\gamma}\left(\hat{\beta}\right) - \gamma \end{pmatrix} &= \left(\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right)\right)^{-1} \\ &\quad \times \hat{\tilde{S}}\left(\hat{\beta}\right)'\left[\left(G^{*} - \hat{G}^{*}\left(\hat{\beta}\right)\right) + \left(\Xi - \hat{\Xi}\left(\hat{\beta}\right)\right)\right], \end{aligned} \tag{3.15}$$

where $\hat{\tilde{S}}\left(\hat{\beta}\right) = \mathcal{Q}_{\hat{\Psi}}S$, $\hat{\Xi}\left(\hat{\beta}\right) = \mathcal{P}_{\hat{\Psi}}\Xi$ and $\hat{G}^{*}\left(\hat{\beta}\right) = \mathcal{P}_{\hat{\Psi}}G^{*}$, with $G^{*} = (g_1^{*}, \ldots, g_N^{*})'$ and $g_i^{*} = z_i g_u\left(\tilde{v}_i\right) + g_\varepsilon\left(\tilde{v}_i\right)$. Define now also $\hat{G}^{*}\left(\beta\right) = \mathcal{P}_{\Psi}G^{*}$, etc.

If the generalized residuals $\tilde{v}_i\left(\beta\right)$ would have been observed, Ψ_{Nk} would be known and the variance of the corresponding estimator, i.e. $\left(\hat{\alpha}_1'\left(\beta\right), \hat{\gamma}'\left(\beta\right)\right) = \left(S'\mathcal{Q}_{\Psi}S\right)^{+}S'\mathcal{Q}_{\Psi}Y$, could have been robustly estimated by the White (1980b) estimator. The asymptotic variance of $\left(\hat{\alpha}_1'\left(\hat{\beta}\right), \hat{\gamma}'\left(\hat{\beta}\right)\right)$ will be shown to be equal to the variance of this hypothetical estimator $\left(\hat{\alpha}_1'\left(\beta\right), \hat{\gamma}'\left(\beta\right)\right)$, pre- and postmultiplied by a correction matrix.

I will now take a closer look at the asymptotic properties of the estimator (3.14). Define $h_x\left(\tilde{v}_i\right)$ and $h_z\left(\tilde{v}_i\right)$ as the conditional expectations

$$\begin{aligned} h_x\left(\tilde{v}_i\right) &= \mathrm{E}\left[x_i \middle| \tilde{v}_i\right] \\ h_z\left(\tilde{v}_i\right) &= \mathrm{E}\left[z_i^{*} \middle| \tilde{v}_i\right], \end{aligned}$$

where it is assumed that the vector-valued functions $h_x\left(\tilde{v}\right)$ and $h_z\left(\tilde{v}\right)$ remain the same

across different observations. We further have that

$$\begin{aligned}\tilde{x}_i &= x_i - h_x(\tilde{v}_i), \\ \tilde{z}_i^* &= z_i^* - h_z(\tilde{v}_i), \\ \tilde{s}_i &= (\tilde{x}_i', \tilde{z}_i^{*\prime})'.\end{aligned}$$

The functions $h_x(\tilde{v})$ and $h_z(\tilde{v})$ are unknown, but they can also be approximated by a series expansion. The convergence rates of $(\hat{\alpha}_1, \hat{\gamma})$ will depend on the approximation error bounds for the functions $g_\varepsilon(\tilde{v})$, $g_u(\tilde{v})$, $h_x(\tilde{v})$ and $h_z(\tilde{v})$ as stated in the following assumption (Donald and Newey (1994)).

Assumption D. There exist functions $e_{g_\varepsilon}(\kappa)$, $e_{g_u}(\kappa)$, $e_{h_x}(\kappa)$ and $e_{h_z}(\kappa)$ such that there are π_ε, π_u, η_x and η_z with

$$\begin{aligned}\sup_{N>0}\left\{N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left\{g_\varepsilon(\tilde{v}_i) - \vec{\varphi}'_{\kappa_\varepsilon}(\tilde{v}_N)\,\pi_\varepsilon\right\}^2\right]\right\}^{\frac{1}{2}} &\leq e_{g_\varepsilon}(\kappa_\varepsilon(N)), \\ \sup_{N>0}\left\{N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left\{z_i\left(g_u(\tilde{v}_i) - \vec{\varphi}'_{\kappa_u}(\tilde{v}_i)\,\pi_u\right)\right\}^2\right]\right\}^{\frac{1}{2}} &\leq e_{g_u}(\kappa_u(N)), \\ \max_j \sup_{N>0}\left\{N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left\{h_{j,x}(\tilde{v}_i) - \vec{\varphi}'_{\kappa_\varepsilon}(\tilde{v}_i)\,\eta_{jx}\right\}^2\right]\right\}^{\frac{1}{2}} &\leq e_{h_x}(\kappa_\varepsilon(N)), \\ \max_k \sup_{N>0}\left\{N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left\{h_{k,z}(\tilde{v}_i) - z_i\vec{\varphi}'_{\kappa_u}(\tilde{v}_i)\,\eta_{k,z}\right\}^2\right]\right\}^{\frac{1}{2}} &\leq e_{h_z}(\kappa_u(N)).\end{aligned}$$

The magnitude of these approximation errors is determined by the degree of smoothness of the functions $g_\varepsilon(\cdot)$, $g_u(\cdot)$, $h_x(\cdot)$ and $h_z(\cdot)$.

Application of the Lyapunov CLT will impose some restrictions on the underlying DGP.

Assumption E. The DGP obeys following assumptions[6]:

1. $\mathrm{Var}\left[y_i \left| x_i, w_i, q_i, \tilde{v}_i\right.\right] \leq \Delta < \infty, \quad \forall i,$

2. $N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[\tilde{s}_i \tilde{s}_i'\right] = \bar{A}_N$ is uniformly positive definite,

3. $N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[\xi_i^2 \tilde{s}_i \tilde{s}_i'\right] = \bar{B}_N$ is uniformly positive definite,

4. $\mathrm{Var}\left[s_i \left| \tilde{v}_i\right.\right] \leq \Delta < \infty, \quad \forall i,$

5. $N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[\left|\tilde{s}_{ji}\xi_i\right|^{2+\delta}\right] \leq \Delta < \infty, \quad \forall j = 1, \ldots, 2K_1 + K_2 + K_3$, for some $\delta > 0$,

6. $N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[\left|\tilde{s}_{ji}\right|^{2+\delta}\right] \leq \Delta < \infty, \quad \forall j = 1, \ldots, 2K_1 + K_2 + K_3$, for some $\delta > 0$.

Condition E.2 is a necessary and sufficient condition for identification of $(\alpha_1', \gamma')'$. If the constant term is not identified, it should be removed from x_i, otherwise E.2 will fail (Donald and Newey (1994)). Conditions E.1 and E.4 guarantee that the unknown functions $g_\varepsilon(\cdot)$, $g_u(\cdot)$, $h_x(\cdot)$ and $h_z(\cdot)$ can be consistently estimated using the series approximations and allow for heteroskedasticity in the residuals ξ_i and $\tilde{s}_{ji}$. Conditions E.2-E.3 and E.5-E.6 ensure the uniform unboundedness of $\bar{A}_N^{-1}$ and $\bar{B}_N^{-1}$, and thus allow application of the Lyapunov CLT to the hypothetical estimator that is obtained when $h_x(\tilde{v})$ and $h_z(\tilde{v})$ would have been known.

$$\begin{pmatrix} \tilde{\alpha}_1 \\ \tilde{\gamma} \end{pmatrix} = \left(\tilde{S}'\tilde{S}\right)^{-1} \tilde{S}' \left(\tilde{S} \begin{pmatrix} \alpha_1 \\ \gamma \end{pmatrix} + \Xi \right).$$

The main results of the chapter are now stated in the following three theorems, the proofs of which are given in Appendix B. They are stated in terms of the order of approximation ($e_{g_\varepsilon}(\kappa)$, $e_{g_u}(\kappa)$, $e_{h_x}(\kappa)$ and $e_{h_z}(\kappa)$) of the unknown functions ($g_\varepsilon(\tilde{v})$, $g_u(\tilde{v})$, $h_x(\tilde{v})$ and $h_z(\tilde{v})$) and in terms of the rate that $\kappa_\varepsilon(N)$ and $\kappa_u(N)$ are allowed to increase with N.

Theorem 3.1. *Given Assumption A-E, and assuming that* $\kappa_\varepsilon(N), \kappa_u(N) = o(N)$*, while still* $\kappa_\varepsilon(N), \kappa_u(N) \to \infty$ *and assuming that* $e_{g_\varepsilon}(\kappa_\varepsilon(N)) \to 0$, $e_{g_u}(\kappa_u(N)) \to 0$, $e_{h_x}(\kappa_\varepsilon(N)) \to$

[6] Let δ denote a small and Δ a large positive constant.

0 *and* $e_{h_z}(\kappa_u(N)) \to 0$, *then*

$$\begin{pmatrix} \hat{\alpha}_1\left(\hat{\beta}\right) - \alpha_1 \\ \hat{\gamma}\left(\hat{\beta}\right) - \gamma \end{pmatrix} = O_p\left(N^{-\frac{1}{2}}\right) + O_p\left(N^{-1}\left(\kappa_\varepsilon + \kappa_u\right)^{\frac{1}{2}}\right) + O_p\left(\left\{e_{h_\varepsilon}(\kappa_\varepsilon) + e_{h_u}(\kappa_u)\right\}\left\{e_{g_\varepsilon}(\kappa_\varepsilon) + e_{g_u}(\kappa_u)\right\}\right) + O_p\left(N^{-\frac{1}{2}} e_{h_x}(\kappa_\varepsilon)\right) + O_p\left(N^{-\frac{1}{2}} e_{h_z}(\kappa_u)\right) + O_p\left(N^{-\frac{1}{2}} e_{g_\varepsilon}(\kappa_\varepsilon)\right) + O_p\left(N^{-\frac{1}{2}} e_{g_u}(\kappa_u)\right).$$

Theorem 3.2. *Given Assumption A-E and the conditions of Theorem 3.1, and assuming that*

$$\sqrt{N}\left(\left\{e_{h_\varepsilon}(\kappa_\varepsilon(N)) + e_{h_u}(\kappa_u(N))\right\}\left\{e_{g_\varepsilon}(\kappa_\varepsilon(N)) + e_{g_u}(\kappa_u(N))\right\}\right) \to 0,$$

then

$$\begin{pmatrix} \sqrt{N}\left(\hat{\alpha}_1\left(\hat{\beta}\right) - \alpha_1\right) \\ \sqrt{N}\left(\hat{\gamma}\left(\hat{\beta}\right) - \gamma\right) \\ \sqrt{N}\left(\hat{\beta} - \beta\right) \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \bar{A}_N^{-1}\left(\bar{B}_N + D\Sigma_{\beta\beta}D'\right)\bar{A}_N^{-1} & \bar{A}_N^{-1}D\Sigma_{\beta\beta} \\ \Sigma_{\beta\beta}D'\bar{A}_N^{-1} & \Sigma_{\beta\beta} \end{pmatrix}$$

with

$$D = E\left[\tilde{s}_i \left.\frac{\partial g^*(x)}{\partial x}\right|_{x=\tilde{v}_i} \frac{\partial \tilde{v}_i}{\partial \beta'}\right].$$

In addition, Σ *can be estimated consistently by substitution of* $\bar{A}_N$, $\bar{B}_N$, $\Sigma_{\beta\beta}$ *and* D *by their respective consistent estimators*

$$\hat{A}_N = N^{-1}\textstyle\sum_{i=1}^N \hat{\tilde{s}}_i\hat{\tilde{s}}_i',$$

$$\hat{B}_N = N^{-1}\textstyle\sum_{i=1}^N \left(\tilde{y}_i - \hat{\tilde{s}}_i'\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\gamma} \end{pmatrix}\right)^2 \hat{\tilde{s}}_i\hat{\tilde{s}}_i',$$

$$\hat{D} = N^{-1}\textstyle\sum_{i=1}^N \hat{\tilde{s}}_i\left(\hat{\beta}\right)\left(\left.\frac{\partial \hat{g}^*(x)}{\partial x}\right|_{x=\hat{\tilde{v}}_i}\right)\left.\frac{\partial \tilde{v}_i}{\partial \beta'}\right|_{\beta=\hat{\beta}},$$

and $\hat{\Sigma}_{\beta\beta}$ *from Assumption C.*

Finally, let the difference between g and $\hat{g}$ be measured by the square root of the sample MSE

$$N^{-\frac{1}{2}} \|\hat{g} - g\| = \sqrt{N^{-1} \sum_{i=1}^{N} \left(\hat{g}\left(\hat{\tilde{v}}_i\right) - g\left(\tilde{v}_i\right)\right)^2},$$

then the following theorem considers its asymptotic behavior.

Theorem 3.3. *Given Assumption A-E and the conditions of Theorem 3.1, it holds that*

$$N^{-\frac{1}{2}} \|\hat{g}^* - g^*\| = O_p\left(\left(\kappa_\varepsilon + \kappa_u\right)^{-\frac{1}{2}} N^{-\frac{1}{2}}\right) + O_p\left(e_{g_\varepsilon}\left(\kappa_\varepsilon\right) + e_{g_u}\left(\kappa_u\right)\right). \tag{3.16}$$

These results provide a cross-section estimator for a broad class of models where unobserved heterogeneity is allowed to interact with a variable.

3.4 Conclusion

In this chapter semiparametric estimation of the outcome equation from the CRC model (Heckman and Vytlacil (1998)) is considered. In the CRC model a random coefficient is correlated with an endogenous regressor. I include an unknown function of the (generalized) residuals (Cox and Snell (1968), Gouriéroux et al. (1987)) in the equation of interest, and propose to estimate the latter semiparametrically, using a series approximation. This methodology is valid in the presence of truncation, sample selection and/or limited dependent variables, resulting in $\sqrt{N}$-consistent, estimation of the parameters of interest.

Chapter 4

Instrumental Variable Estimation of the Correlated Random Coefficient Model with Panel Data

Abstract. This paper considers some estimation methods for the Correlated Random Coefficient (CRC) model when panel data are available, which allows for estimation of the average treatment effect by IV as described by Wooldridge (2003a), using Hausman and Taylor (1981) style instruments.

4.1 Introduction

For a considerable time, different estimation procedures have been proposed for the Correlated Random Coefficient model, albeit not always under this name. Whenever a choice is an optimal action, instead of an exogenous datum, the subsample of individuals choosing a particular option is nonrandom. "Bias arises because maximization truncates the underlying disturbances of the equations" (Garen (1984)). More strongly, it has been demonstrated that the optimality of a choice can imply its correlation with its effect (see Garen (1984) and Card (2001) in the context of returns to schooling). Different procedures have been suggested to estimate the mean return to such a choice variable, called the average treatment effect (ATE).

Garen (1984) assumes that the decision of interest is continuous and is solely determined by exogenous variables in a secondary equation. This author proposes an alternative estimator of the primary equation that includes both the residual of the secondary equation

and the cross-product of this residual with the choice variable. This two-step procedure is termed the selectivity bias (SB) method. Heckman and Vytlacil (1998) propose a two-stage plug-in estimator using the product of the predicted values of the choice variable with the observed exogenous determinants of its return. Wooldridge (2003a) shows that a standard instrumental variable estimator where the set of instruments consists of the cross-products of the instruments of the choice variable with the exogenous determinants of its return, consistently estimates the ATE. Wooldridge (2003b) discusses the robustness of the fixed effects (FE) and the first difference (FD) estimators for estimation of the ATE in CRC panel data models with individual-specific slopes. Belzil and Hansen (2002) conduct a structural analysis of the CRC wage regression model with individual-specific returns to schooling and experience.

On the other hand, unobservable individual- and period-specific effects can be controlled for with panel data. In increasing order of efficiency and strength of assumptions, Hausman and Taylor (HT) (1981), Amemiya and MaCurdy (AM) (1986) and Breusch, Mizon and Schmidt (BMS) (1989) proposed instrumental variable estimation of a linear static panel data model when individual effects are correlated with a subset of the regressors. These respective estimators are compared using a returns to schooling example by Cornwell and Rupert (1988) and by Baltagi and Khanti-Akom (1990). Wyhowski (1994) presents estimators for generalized HT, AM and BMS models with two-way error components.

Random coefficient models, finally, are widely studied, particularly since the late sixties (Rao (1965)), mostly from the point of view of efficiency (Hildreth and Houck (1968)). Swamy (1970) and Hsiao (1975, 1974) propose estimators for the linear model with random coefficients adapted to panel data with cross-sectional, time effects or both.

In this chapter a CRC model for linear static panel data is considered, where each random slope can be written as a linear function of some variables and an additive error term that consists of three components: an individual-specific, a period-specific and an idiosyncratic component. In addition, these random slopes are allowed to be correlated with some regressors. Assuming conditional constancy of covariances (as Heckman and Vytlacil (1998) and Wooldridge (2003a)), I first propose to estimate the ATE by standard IV methods, using as instruments cross-products of the set of HT instruments of the model variables and the set of HT instruments of the slope variables, after removal of product variables that occur more than once. In addition, the properties of the Aitken IV and feasible Aitken IV estimators (Mandy and Martins-Filho (1994))) are studied.

The proposed estimators extend existing estimation methods in several ways. First of

all Woolridge's (2003a) IV estimator for a CRC model in cross-sections is extended to panel data, whereby the specification of the correlation between regressors and random coefficients is kept as broad as possible. Secondly, as far as the constant term is also associated with a random and correlated coefficient, the two-way unobservable effects methodology of Wyhowski (1994) is extended in such a way that also correlation between some variables and the idiosyncratic error terms is allowed. Finally, the proposed estimator generalizes Hsiao's (1975) random coefficient estimator in the sense that some degree of correlation between coefficients and variables is allowed and in the sense that the coefficients are also allowed to exhibit idiosyncratic variation.

The organization of the chapter is as follows. Section 4.2 presents the basic framework. Section 4.3 discusses identification of the model. In section 4.4 estimation is considered. Finally, section 4.5 concludes and proofs are given in Appendix C.

4.2 Framework and Preliminaries

Consider the following model

$$y_{it} = (1, x'_{it})\,\alpha_{it}, \tag{4.1}$$

where y_{it} is the outcome for individual i at time t, with $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where x_{it} is a $K \times 1$ vector of variables with associated 'parameter' vector, $\alpha'_{it} = \left(\alpha_{0,it}, \alpha'_{1,it}\right)$, where $\alpha_{0,it}$ corresponds to the 'constant term'. Suppose now that vector α'_{it} can be written as

$$\alpha_{it} = \begin{pmatrix} \mu_{\alpha_0} & \gamma'_{0X} & \gamma'_{0S} \\ \mu_{\alpha_1} & \Gamma_{XX} & \Gamma_{XS} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ x_{it} - \mu_X \\ s_{it} - \mu_S \end{pmatrix} + u_{it}, \tag{4.2}$$

where s_{it} is a $J \times 1$ vector of variables, all distinct from the variables in x_{it}, with $\mu_X = \mathrm{E}\left[x_{it}\right]$ and $\mu_S = \mathrm{E}\left[s_{it}\right]$. I restrict the parameter matrix in the following ways. It holds that

$$\mu_\alpha \;=\; \left(\mu_{\alpha_0}, \mu'_{\alpha_1}\right)' \;=\; \mathrm{E}\left[\alpha_{it}\right] \;<\; \infty,$$

with μ_{α_1} the parameters of interest. In addition, γ_{0S} is a $J \times 1$ parameter vector, $\gamma_{0X} = \mathbf{O}_{K\times 1}$ a $K \times 1$ vector identical to zero[1], Γ_{XS} is a $K \times J$ parameter matrix and Γ_{XX} is a lower triangular $K \times K$ matrix. The restrictions on the elements of Γ are needed for identification purposes. $\gamma_{0X} = \mathbf{O}_{K\times 1}$, for instance, means that $\alpha_{0,it}$ does not depend on

[1] In general $\mathbf{O}_{L\times M}$ denotes a $L \times M$ matrix with all elements equal to zero.

x_{it}. There is thus only one term in x_{it} that appears after substitution of (4.2) in (4.1).

Specification (4.2) allows the random coefficients to be composed of a term dependent of observed covariates and a random error vector as in Heckman and Vytlacil (1998). The following assumption is made about the error vector $u_{it} = \left(u_{0,it}, u'_{1,it}\right)'$ in the specification (4.2) of the random coefficients.

Assumption A. The $(K+1) \times 1$ error vector u_{it} from (4.2) can be written as the sum of three independent components

$$u_{it} = \lambda_i + \tau_t + \xi_{it}, \tag{4.3}$$

which are distributed as

$$\begin{pmatrix} \lambda_i \\ \tau_t \\ \xi_{it} \end{pmatrix} \sim \text{IID}\left(0; \begin{pmatrix} \Sigma_\lambda & 0 & 0 \\ 0 & \Sigma_\tau & 0 \\ 0 & 0 & \Sigma_\xi \end{pmatrix}\right). \tag{4.4}$$

Since λ_i, τ_t and ξ_{it} are thus far $(K+1) \times 1$ vectors, we have that Σ_λ, Σ_τ and Σ_ξ are $(K+1)\times(K+1)$ matrices. But, in general, these vectors need not be of equal length, since some components of u_{it} may be equal to zero for some variables. The corresponding rows and columns need than to be removed from (4.4). Making abstraction of the deterministic component in (4.2), the difference between (4.3)-(4.4) and Hsiao's (1975) specification, is that, in the notation above, he imposes the restriction $\xi_{1,it} = 0$. In case of the usual two-way error components model (Wyhowski (1994)), i.e. only the constant term is a random coefficient, Σ_λ, Σ_τ and Σ_ξ reduce to the scalars σ^2_λ, σ^2_τ and σ^2_ξ. Remark that no assumption is made so far about the exogeneity of x_{it} and s_{it}.

Using (4.2), equation (4.1) can be rewritten as

$$y_{it} = (1, x'_{it}) \begin{pmatrix} \mu_{\alpha_0} & \gamma'_{0X} & \gamma'_{0S} \\ \mu_{\alpha_1} & \Gamma_{XX} & \Gamma_{XS} \end{pmatrix} \begin{pmatrix} 1 \\ x_{it} - \mu_X \\ s_{it} - \mu_S \end{pmatrix} + u_{0,it} + x'_{it} u_{1,it}$$

or[2]

$$y_{it} = r'_{it} \begin{pmatrix} \mu_\alpha \\ \gamma_{0S} \\ \text{Vech}\left[\Gamma_{XX}\right] \\ \text{Vec}\left[\Gamma_{XS}\right] \end{pmatrix} + z'_{it} u_{it}, \tag{4.5}$$

with

$$r_{it} = \left(z'_{it}, \left(s_{it} - \mu_S\right)', \text{Vech}\left[x_{it} \cdot \left(x_{it} - \mu_X\right)'\right]', \text{Vec}\left[x_{it} \cdot \left(s_{it} - \mu_S\right)'\right]'\right)'$$

and $z_{it} = (1, x'_{it})'$. Remark that the only difference between x_{it} and s_{it} is that the former vector is multiplicatively present in the error term, while the latter is not[3].

The following conventions apply in the rest of the chapter: let x_{it} denote the it-th observation of the $(K \times 1)$ vector x, then we have that

$$\underset{(T\times K)}{X_j} = \begin{pmatrix} x'_{j1} \\ x'_{j2} \\ \vdots \\ x'_{jT} \end{pmatrix}, \qquad \underset{(NT\times K)}{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix},$$

$$\underset{(NT\times KT)}{X^{(T)}} = \begin{pmatrix} x'_{11} & 0 & \cdots & 0 \\ 0 & x'_{12} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x'_{1T} \\ \vdots & \vdots & \vdots & \vdots \\ x'_{N1} & 0 & \cdots & 0 \\ 0 & x'_{N2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x'_{NT} \end{pmatrix}, \quad \underset{(KNT\times 1)}{X^{(S)}} = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1T} \\ \vdots \\ x_{N1} \\ x_{N2} \\ \vdots \\ x_{NT} \end{pmatrix},$$

[2]The matrix operation $\text{Vec}(X)$ creates a $JK \times 1$ column vector by stacking the columns of the $J \times K$ matrix X on top of each other. Likewise, $\text{Vech}(Y)$ creates a $J(J+1)/2 \times 1$ column vector by stacking the columns of the $J \times J$ matrix Y, starting at the diagonal elements.

[3]Remark that I do not allow all components of u_{it} to be zero for a particular variable. In that case it would be no longer multiplicatively present in the error term and should be moved from x_{it} to s_{it}.

$$\underset{(NT\times KN)}{X^{(B)}} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_N \end{pmatrix},$$

$$\underset{(NT\times KNT)}{X^{(D)}} = \begin{pmatrix} x'_{11} & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & x'_{12} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x'_{1T} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & x'_{N1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & x'_{N2} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & x'_{NT} \end{pmatrix}.$$

These conventions will come in handy later on.

Since I am looking to estimate (4.5) with Hausman and Taylor (1981)-like instruments, the following definitions will be needed. Let $\mathcal{P}_A$ be the projection onto the column space of the matrix A, so that $\mathcal{P}_A = A\left(A'A\right)^{+}A'$, where the $+$ superscript denotes the Moore-Penrose generalized inverse. Then $\mathcal{Q}_A = I - \mathcal{P}_A$ is the projection operator onto the null space of A. Letting ι_L be a $(L \times 1)$ vector of ones and $J_L = \iota_L \iota'_L$ a $(L \times L)$ matrix of ones, we can define

$$\mathcal{P}_{\iota_T} = T^{-1}J_T, \quad \mathcal{Q}_{\iota_T} = I_T - \mathcal{P}_{\iota_T}; \qquad \mathcal{P}_{\iota_N} = N^{-1}J_N, \quad \mathcal{Q}_{\iota_N} = I_N - \mathcal{P}_{\iota_N}.$$

Now define the matrix of individual dummy variables $D_i = I_N \otimes \iota_T$ and the matrix of period dummy variables $D_p = \iota_N \otimes I_T$, then we have the following four projection matrices

$$\begin{aligned} \mathcal{P}_{D_i} &= T^{-1}I_N \otimes J_T = I_N \otimes \mathcal{P}_{\iota_T}, \quad \mathcal{Q}_{D_i} = I_{NT} - \mathcal{P}_{D_i} = I_N \otimes \mathcal{Q}_{\iota_T}, \\ \mathcal{P}_{D_p} &= N^{-1}J_N \otimes I_T = \mathcal{P}_{\iota_N} \otimes I_T, \quad \mathcal{Q}_{D_p} = I_{NT} - \mathcal{P}_{D_p} = \mathcal{Q}_{\iota_N} \otimes I_T. \end{aligned}$$

$\mathcal{P}_{D_i}$ transforms the data into a vector of group means, $\mathcal{Q}_{D_i}$ produces deviations from group means, $\mathcal{P}_{D_p}$ transforms the data into a vector of period means and $\mathcal{Q}_{D_p}$ produces deviations

from period means. Combining both orthogonal projections, we have that

$$X = \mathcal{P}_m X + \mathcal{P}_i X + \mathcal{P}_p X + \mathcal{P}_r X,$$

with

$$\begin{aligned} \mathcal{P}_m &= \mathcal{P}_{D_i}\mathcal{P}_{D_p}, \\ \mathcal{P}_i &= \mathcal{P}_{D_i}\mathcal{Q}_{D_p}, \\ \mathcal{P}_p &= \mathcal{Q}_{D_i}\mathcal{P}_{D_p}, \\ \mathcal{P}_r &= \mathcal{Q}_{D_i}\mathcal{Q}_{D_p}, \end{aligned} \tag{4.6}$$

where all four terms are orthogonal to each other. These orthogonal projections allow "the isolation of any potential correlation between error components and regressors to the associated orthogonal subspace" (Wyhowski (1994)).

4.3 Identification

Suppose now that we possess prior information[4] that allows us to discern three distinct partitions of the vector x_{it}. The first partition distinguishes between the components of x_{it} along the individual-specific dimension.

Assumption B.1. X can be partitioned as

$$X = \left(X^a, X^{\dot{a}}, X^{\ddot{a}}\right),$$

such that

$$\begin{aligned} \mathrm{E}\left[\lambda_i \left| x_{jt}^a \right.\right] &= 0, \qquad \forall i,j \quad \text{and} \quad \mathcal{P}_i X^a \neq 0, \\ \mathrm{E}\left[\lambda_i \left| x_{jt}^{\dot{a}} \right.\right] &= \begin{cases} \mu_{\lambda|x} \neq 0, & i = j \\ 0, & i \neq j \end{cases} \quad \text{and} \quad \mathcal{P}_i X^{\dot{a}} \neq 0, \\ \mathcal{P}_i X^{\ddot{a}} &= 0. \end{aligned}$$

The first subset consists of k_a variables with a non-zero individual-specific component, of which λ_i is mean-independent, the second subset contains $k_{\dot{a}}$ variables with a non-

[4] As is done in HT, AM, BMS and Wyhowski (1994).

zero individual-specific component, that are correlated with λ_i and in the third subset are situated $k_{\ddot{a}} = K - k_a - k_{\dot{a}}$ variables without an individual-specific component. Analogously, two more partitions are possible, one considers the period-specific part of x_{it}, the other one the idiosyncratic part of x_{it}.

Assumption B.2. X can be partitioned as

$$X = \left(X^b, X^{\dot{b}}, X^{\ddot{b}}\right)$$

where

$$\begin{aligned} \mathrm{E}\left[\tau_s \left| x_{it}^b \right.\right] &= 0, \qquad \forall s,t \quad \text{and} \quad \mathcal{P}_p X^b \neq 0, \\ \mathrm{E}\left[\tau_s \left| x_{it}^{\dot{b}} \right.\right] &= \begin{cases} \mu_{\tau|x} \neq 0, & s = t \\ 0, & s \neq t \end{cases} \quad \text{and} \quad \mathcal{P}_p X^{\dot{b}} \neq 0, \\ \mathcal{P}_p X^{\ddot{b}} &= 0. \end{aligned}$$

Assumption B.3. X can be partitioned as

$$X = \left(X^c, X^{\dot{c}}, X^{\ddot{c}}\right)$$

with

$$\begin{aligned} \mathrm{E}\left[\xi_{is} \left| x_{jt}^c \right.\right] &= 0, \qquad \forall i,j,s,t \quad \text{and} \quad \mathcal{P}_r X^c \neq 0, \\ \mathrm{E}\left[\xi_{is} \left| x_{jt}^{\dot{c}} \right.\right] &= \begin{cases} \mu_{\xi|x} \neq 0, & i = j \wedge s = t \\ 0, & i \neq j \vee s \neq t \end{cases} \quad \text{and} \quad \mathcal{P}_r X^{\dot{c}} \neq 0, \\ \mathcal{P}_r X^{\ddot{c}} &= 0. \end{aligned}$$

The variables that are correlated with τ_t, respectively ξ_{it}, are grouped into the subset $X^{\dot{b}}$, respectively $X^{\dot{c}}$. The subset of variables with zero time-specific, respectively idiosyncratic, component is given by $X_{it}^{\ddot{b}}$, respectively $X_{it}^{\ddot{c}}$.

Considered together, these three partitions define an overall partition of X with $3^3 - 1$ classes

$$X = \left(X^{abc}, X^{\dot{a}bc}, X^{\ddot{a}bc}, \ldots, X^{a\dot{b}\ddot{c}}, X^{\dot{a}\dot{b}\ddot{c}}\right).$$

Remark that $X^{\ddot{a}\dot{b}\ddot{c}}$ is not included since it is proportional to the constant term. The subset

$X^{a\dot{b}\ddot{c}}$, for instance, consists of $k_{a\dot{b}\ddot{c}}$ variables with zero idiosyncratic component and non-zero individual- and time-specific components of which λ_i is mean-independent, but with which τ_t is correlated. Finally, when only two of the three partitions are considered the possible subsets are denoted with doubly superscripted letters. The matrix $X^{\dot{b}\ddot{c}}$, for instance, is equal to $\left(\begin{array}{ccc} X^{a\dot{b}\ddot{c}}, & X^{\dot{a}\dot{b}\ddot{c}}, & X^{\ddot{a}\dot{b}\ddot{c}} \end{array} \right)$ and has $k_{\dot{b}\ddot{c}} = k_{a\dot{b}\ddot{c}} + k_{\dot{a}\dot{b}\ddot{c}} + k_{\ddot{a}\dot{b}\ddot{c}}$ columns.

In a similar way the vector s_{it} can be partitioned in 26 classes as

$$S = \left(S^{abc}, S^{\dot{a}bc}, S^{\ddot{a}bc}, \ldots, S^{a\ddot{b}\ddot{c}}, S^{\dot{a}\ddot{b}\ddot{c}} \right),$$

where the number of variables in S^{abc} is given by j_{abc}, the number in $S^{\dot{a}bc}$ by $j_{\dot{a}bc}$, etc. Defining the matrices V_X and V_S, with $k_{V_X} = k_a + k_b + k_c$ elements and $j_{V_S} = j_a + j_b + j_c$ elements, respectively, as

$$\begin{aligned} V_X &= \left(\mathcal{P}_i X^a, \mathcal{P}_p X^b, \mathcal{P}_r X^c \right), \\ V_S &= \left(\mathcal{P}_i S^a, \mathcal{P}_p S^b, \mathcal{P}_r S^c \right), \end{aligned}$$

it holds that

$$\begin{aligned} \operatorname*{plim}_{N,T\to\infty} \frac{(V_X, V_S)' U}{NT} &= \operatorname*{plim}_{N,T\to\infty} \frac{(V_X, V_S)' (\Lambda + \Upsilon + \Xi)}{NT} \\ &= 0. \end{aligned} \tag{4.7}$$

Equation (4.7) will now be verified term by term. First, it holds that $\operatorname*{plim}_{NT\to\infty} \frac{V_X'\Lambda}{NT} = 0$, since

- $\mathrm{E}\left[x_{it}^a \lambda_i'\right] = 0$, by Assumption B.1,
- $X^{b\prime}\mathcal{P}_p'\Lambda = X^{b\prime}\mathcal{P}_p'\mathcal{P}_i\Lambda = 0$, and
- $X^{c\prime}\mathcal{P}_r'\Lambda = X^{c\prime}\mathcal{P}_r'\mathcal{P}_i\Lambda = 0$.

Secondly, $\operatorname{plim}_{N,T\to\infty} \frac{V_X'\Upsilon}{NT} = 0$ is verified similarly. Finally, we have that $\operatorname{plim}_{N,T\to\infty} \frac{V_X'\Xi}{NT} = 0$, because

- $\text{plim}_{N,T\to\infty}\frac{X^{a\prime}\mathcal{P}_i'\Xi}{NT}=0$, since it holds that

$$\begin{aligned}\text{plim}_{N,T\to\infty}\frac{X^{a\prime}\mathcal{P}_i'\Xi}{NT} &= \text{plim}_{N,T\to\infty}(NT)^{-1}\textstyle\sum_{i,t}\left(T^{-1}\sum_s x_{is}^a-(NT)^{-1}\sum_{j,s}x_{js}^a\right)\xi_{it}\\ &= \text{plim}_{N,T\to\infty}N^{-1}\textstyle\sum_i\left(T^{-1}\sum_s x_{is}^a\right)\left(T^{-1}\sum_t\xi_{it}\right)\\ &\quad -\text{plim}_{N,T\to\infty}(NT)^{-2}\left(\textstyle\sum_{j,s}x_{js}^a\right)\left(\sum_{i,t}\xi_{it}\right)\\ &= T^{-1}\text{E}\left[x_{it}^{a\dot{c}}\xi_{it}'\right]-(NT)^{-1}\text{E}\left[x_{it}^{a\dot{c}}\xi_{it}'\right]\\ &= O\left(T^{-1}\right).\end{aligned}$$

- $\text{plim}_{N,T\to\infty}\frac{X^{b\prime}\mathcal{P}_p'\Xi}{NT}=0$, by a similar proof,
- $\text{plim}_{N,T\to\infty}\frac{X^{c\prime}\mathcal{P}_r'\Xi}{NT}=0$, by Assumption B.3.

A similar reasoning proves $\text{plim}_{N,T\to\infty}\frac{V_S'U}{NT}=0$.

Reconsidering expression (4.5), I assume the following about the correlation between x_{it} and u_{it}.

Assumption C. *Conditional constancy of covariances* (Heckman and Vytlacil (1998), Wooldridge (2003a)), i.e.

$$\begin{aligned}\text{E}\left[x_{it}'u_{1;it}\left|v_{X,it},v_{S,it}\right.\right] &= \text{E}\left[x_{it}'\lambda_{1;i}\left|v_{X,it},v_{S,it}\right.\right]+\text{E}\left[x_{it}'\tau_{1;t}\left|v_{X,it},v_{S,it}\right.\right]\\ &\quad +\text{E}\left[x_{it}'\xi_{1;it}\left|v_{X,it},v_{S,it}\right.\right]\\ &= \sigma_{x\lambda}+\sigma_{x\tau}+\sigma_{x\xi}=\sigma_{xu},\end{aligned}\tag{4.8}$$

where neither $\sigma_{x\lambda}$, $\sigma_{x\tau}$ or $\sigma_{x\xi}$ is a function of $v_{X,it}$ nor of $v_{S,it}$.

Taking into account Assumption C, (4.5) can be rewritten as

$$Y = R\varphi+H,\tag{4.9}$$

with $\varphi=\left(\mu_{\alpha_0}+\sigma_{xu},\mu_{\alpha_1}',\gamma_{0S}',\text{Vech}\left[\Gamma_{XX}\right]',\text{Vec}\left[\Gamma_{XS}\right]'\right)'$ and $H=H_\lambda+H_\tau+H_\xi$, where $H_\lambda=Z^{(D)}\Lambda^{(S)}-\sigma_{x\lambda}\iota_{NT}$, $H_\tau=Z^{(D)}\Upsilon^{(S)}-\sigma_{x\tau}\iota_{NT}$ and $H_\xi=Z^{(D)}\Xi^{(S)}-\sigma_{x\xi}\iota_{NT}$, recalling that $Z^{(D)}=(\iota_{NT},X)^{(D)}$, since $z_{it}=(1,x_{it}')'$.

Define now the $k_V\equiv(1+k_{V_X})(1+k_{V_X}/2+j_{V_S})$-length column vector v_{it} of instruments as the vector that results if we cross-multiply $v_{X,it}$ with $v_{S,it}$, and then remove

collinear elements

$$v_{it} = \left(1, v'_{X,it}, v'_{S,it}, \text{Vech}\left[v_{X,it} \cdot v'_{X,it}\right]', \text{Vec}\left[v_{X,it} \cdot v'_{S,it}\right]'\right)'.$$

Remark that the removal of collinear elements corresponds to the restrictions imposed on the parameter matrix in (4.2). I make the following assumption about the vector v_{it}.

Assumption D. It holds that $\text{rank}\left[V'V\right] = k_V$.

The following proposition concerns the identification of model (4.9).

Proposition 4.1. *Taking into account Assumption A-D, sufficient conditions for the identification of φ in (4.9),* except μ_{α_0}*, are given by*

$$\begin{aligned} m_{\ddot{b}\ddot{c}} &\leq m_a, \\ m_{\ddot{a}\ddot{c}} &\leq m_b, \\ m_{\ddot{a}\ddot{b}} &\leq m_c, \end{aligned} \tag{4.10}$$

$$\begin{aligned} m_{\ddot{a}} &\leq m_b + m_c, \\ m_{\ddot{b}} &\leq m_a + m_c, \\ m_{\ddot{c}} &\leq m_a + m_b, \end{aligned} \tag{4.11}$$

and

$$\begin{aligned} & m_{\dot{a}\dot{b}\dot{c}} + (m_{\ddot{a}\dot{b}\dot{c}} + m_{\dot{a}\ddot{b}\dot{c}} + m_{\dot{a}\dot{b}\ddot{c}}) + (m_{\dot{a}\ddot{b}\ddot{c}} + m_{\ddot{a}\dot{b}\ddot{c}} + m_{\ddot{a}\ddot{b}\dot{c}}) \\ \leq\ & 2m_{abc} + (m_{\dot{a}bc} + m_{a\dot{b}c} + m_{ab\dot{c}}) + (m_{\ddot{a}bc} + m_{a\ddot{b}c} + m_{ab\ddot{c}}), \end{aligned} \tag{4.12}$$

for $m = j, k$.

A proof of Proposition 4.1 is given in Appendix C. Conditions (4.10) state that the number of valid instruments for X, respectively S, in a particular dimension needs to be at least as high as the number of variables in X, respectively S, that have only one component in that particular direction. For example $k_{\ddot{b}\ddot{c}} \leq k_a$, demands that the number of individual-specific variables in X, i.e. $X^{\ddot{b}\ddot{c}}$, is not larger than the number of variables that provide individual-specific instruments, i.e. X^a. Conditions (4.11) ensure that variables in X, respectively S, with one component missing are identified. For instance, the $j_{\ddot{a}}$ variables

$X^{\ddot{a}}$ that miss the individual-specific component can only have time-specific or idiosyncratic (parts of) variables as instruments, of which the total number is $j_b + j_c$. Condition (4.12), finally, states that the number of variables generating no instrument should not be greater than the number of variables generating two instruments plus two times the number of exogenous variables. Define the surplus of instruments for a variable as the number of valid instruments we can generate from this variable, minus one. For example, if X_1 is a variable with every component exogenous, i.e. it is a column of X^{abc}, then from this variable spring three valid instruments, namely $\mathcal{P}_i X_1$, $\mathcal{P}_p X_1$ and $\mathcal{P}_r X_1$. One is used to identify X_1 itself, but the two other can serve as instruments for other variables, hence we have a surplus of two instruments. Similarly, a variable with all components endogenous, has a surplus of minus one. Condition (4.12) can now states that the total surplus should at least be zero. In other words, condition (4.12) ensures that the number of valid instruments generated from X, respectively S, is larger than the number of variables in X, respectively S.

In the framework outlined above, in which $\alpha_{0,it}$ is not a function of $(x_{it} - \mu_X)$ and Γ_{XX} is a lower triangular matrix and under Assumption A-D, the parameter vector of interest, μ_{α_1}, and all other components of φ, *except the constant term*, are identified, provided that the conditions from Proposition 4.1 hold. Remark that Proposition 4.1 ensures that

$$(1+K)\left(1+\frac{K}{2}+J\right) \;\leq\; (1+k_{V_X})\left(1+\frac{k_{V_X}}{2}+k_{V_S}\right),$$

i.e. the number of instruments is not smaller than the number of variables in model (4.5). Identification of the constant term is only ensured if Assumption C is strengthened to $\sigma_{xu} = 0$.

The model described so far in this chapter encompasses some well-known models in the literature. It simplifies to a generalization of Hsiao's (1974) random coefficient model, when (4.2) simplifies to $\left(\alpha'_{0,it}, \alpha'_{1,it}\right) = \left(\mu'_{\alpha_0}, \mu'_{\alpha_1}\right) + u'_{it}$ and when $\mathrm{E}\left[u_{it} \left| x_{it}\right.\right] = 0$. If (4.1)-(4.2) is reduced to $y_{it} = \alpha_{0,it} = \mu_{\alpha_0} + \Gamma_{XS}\left(s_{it} - \mu_S\right) + u_{it}$, (4.5) collapses into a slight generalization of Wyhowski's (1994) two-way error component model.

4.4 Estimation

4.4.1 IV Estimation

A first, obvious, way to estimate (4.9) is simple IV estimation, using V as instruments, which results in

$$\hat{\varphi}_{IV} = \left(R'\mathcal{P}_V R\right)^{-1} R'\mathcal{P}_V Y. \tag{4.13}$$

I will now derive the convergence rate and asymptotic distribution of this estimator. It is important to note in this respect that the order of convergence is not necessarily identical for all elements of $\hat{\varphi}_{IV}$. This fact, apparent in work on error components (see, for example, Wyhowski (1994)), remains rather obscure in the literature on random coefficients.

The variables in R can be partitioned in the same 26 partition classes as the variables of X and S earlier on. Each partition class can be easily constructed from the classes of X and S. The subset of individual-specific variables, $R^{\ddot{b}\ddot{c}}$, for instance, consists of $x_{it}^{\ddot{b}\ddot{c}}$, $s_{it}^{\ddot{b}\ddot{c}} - \mu_{S^{\ddot{b}\ddot{c}}}$, $\text{Vech}\left[x_{it}^{\ddot{b}\ddot{c}} \cdot \left(x_{it}^{\ddot{b}\ddot{c}} - \mu_{X^{\ddot{b}\ddot{c}}}\right)'\right]$ and $\text{Vec}\left[x_{it}^{\ddot{b}\ddot{c}} \cdot \left(s_{it}^{\ddot{b}\ddot{c}} - \mu_{S^{\ddot{b}\ddot{c}}}\right)\right]$. All variables are now grouped into three subsets within which the order of convergence is identical for all variables.

- R_i contains at least the variables which are individual-specific, i.e. $R^{\ddot{b}\ddot{c}}$,
- R_p consists at least of the period specific variables, i.e. $R^{\ddot{a}\ddot{c}}$, and
- R_r contains at least the 'pure' idiosyncratic variables, $R^{\ddot{a}\ddot{b}}$.

It is possible that the convergence rates differ between the different subsets R_i, R_p and R_r. The exact composition of each group is made explicit in Theorem 4.1 below. Define now

$$\begin{aligned} P &= \operatorname*{plim}_{N,T\to\infty} \left[(NT)^{-1} R'\mathcal{P}_V R\right], \\ P_j &= \operatorname*{plim}_{N,T\to\infty} \left[(NT)^{-1} R_j'\mathcal{P}_{V_j} R_j\right] \quad \text{for } j = i,p,r, \\ V_j &= \mathcal{P}_j V \qquad\qquad\qquad\qquad\qquad\qquad \text{for } j = i,p,r. \end{aligned}$$

The following theorem concerns the asymptotic distribution of $\hat{\varphi}_{IV}$.

Theorem 4.1. *Given Assumption A-D, given the sufficient identification conditions from Proposition 4.1, and assuming $H = H_\lambda + H_\tau + H_\xi$, $\hat{\varphi}_{IV}$ is consistent and $\Psi\left(\hat{\varphi}_{IV} - \varphi\right) \xrightarrow{d}$*

$N\left(0; P^{-1}\Gamma P^{-1}\right)$ *as* $N, T \to \infty$, *where*

$$\begin{aligned}
\Gamma &= \lim_{N,T\to\infty} E\left[(NT)^{-2}\,\Psi R'\mathcal{P}_V\Omega_\eta\mathcal{P}_V R\Psi\right], \\
\Psi &= Diag_{j=i,p,r}\left[\sqrt{M_j}I_{\dim R_j}\right], \\
R &= (R_i, R_p, R_r) \\
\Omega_\eta &= Var[H].
\end{aligned}$$

With respect to the grouping of variables into the sets R_i, R_p *and* R_r, *we discern the following mutually exclusive cases:*

I. If $N = O(T)$, $R_i = \left(R^a, R^{\dot{a}}\right)$, $R_p = \left(R^{\ddot{a}b}, R^{\ddot{a}\dot{b}}\right)$, *and* $R_r = R^{\ddot{a}\ddot{b}}$;

II. If $T = o(N)$, $R_i = \left(R^{a\ddot{b}}, R^{\dot{a}\ddot{b}}\right)$, $R_p = \left(R^b, R^{\dot{b}}\right)$ *and* $R_r = R^{\ddot{a}\ddot{b}}$.

With respect to the consistency rates, one of the four following, mutually exclusive cases apply:

1. *If* $\mathcal{P}_i H = \mathcal{P}_i(H_\lambda + H_\xi)$ *and* $\mathcal{P}_p H = \mathcal{P}_p(H_\tau + H_\xi)$, *then it holds that* $M_i = N$, $M_p = T$ *and* $M_r = NT$. *If* $Z^{(D)} = I_{NT}$, Γ *can simplified to* $Diag\left[\sigma_\lambda^2 P_i, \sigma_\tau^2 P_p, \sigma_\xi^2 P_r\right]$;

2. *If* $\mathcal{P}_i H = \mathcal{P}_i(H_\lambda + H_\xi)$ *and* $\mathcal{P}_p H \neq \mathcal{P}_p(H_\tau + H_\xi)$, *then* $M_i = M_r = N$ *and* $M_p = \min(N, T)$;

3. *If* $\mathcal{P}_i H \neq \mathcal{P}_i(H_\lambda + H_\xi)$ *and* $\mathcal{P}_p H = \mathcal{P}_p(H_\tau + H_\xi)$, *then* $M_i = \min(N, T)$ *and* $M_p = M_r = T$;

4. *If* $\mathcal{P}_i H \neq \mathcal{P}_i(H_\lambda + H_\xi)$ *and* $\mathcal{P}_p H \neq \mathcal{P}_p(H_\tau + H_\xi)$, *then* $M_i = M_p = M_r = \min(N, T)$.

Remark 2. Case I encompasses $N \sim T$, in which case the distinction between R_i, R_p and R_r is unnecessary, since the consistency rates are identical in all cases 1-4. For consistency with their definition, the distinct notation of these subsets is maintained and the subsets R^{ab}, $R^{\dot{a}b}$, $R^{a\dot{b}}$ and $R^{\dot{a}\dot{b}}$ of R were arbitrarily allocated to R_i, which makes it identical to case I.

Taking into account (4.3) and (4.9), the condition $\mathcal{P}_p H = \mathcal{P}_p(H_\tau + H_\xi)$ verifies the presence of individual-specific components in the error term H. It is satisfied if the only coefficients that show individual-specific variation are the ones associated with individual-specific variables or with the constant term. In that case it holds that $\mathcal{P}_p H_\lambda = \mathcal{P}_r H_\lambda = 0$,

a fact that insulates the individual-specific part of u_{it} from the period-specific and the idiosyncratic variables. Similarly, condition $\mathcal{P}_i H = \mathcal{P}_i (H_\lambda + H_\xi)$ verifies the presence of period-specific components in the error term H and it is fulfilled if coefficients exhibiting period-specific variation belong solely to period-specific variables or to the constant term. The combination of both conditions occurs in a model where the only random coefficient is associated with the constant term, i.e. when the considered model collapses into a two-way error component model. As a general principle it holds that parameters of variables with more than one component belong to the group with the *lowest* convergence rate.

4.4.2 Aitken Estimation

If Ω_η is known, the efficiency of $\hat{\varphi}_{IV}$ can be improved upon by an IV version of the Aitken estimator (Mandy and Martins-Filho (1994))

$$\hat{\varphi}_{IVA} = \left(R' \mathcal{P}_V \Omega_\eta^{-1} R\right)^{-1} R' \mathcal{P}_V \Omega_\eta^{-1} Y. \tag{4.14}$$

Defining $P_\Omega = \text{plim}_{N,T\to\infty} \left[(NT)^{-1} R' \mathcal{P}_V \Omega_\eta^{-1} R\right]$ and R_i, R_p and R_r as in Theorem 4.1, the asymptotic distribution of $\hat{\varphi}_{IVA}$ is established in the following theorem, which is also proved in Appendix C.

Theorem 4.2. *Under Assumption A-D, and the sufficient identification conditions from Proposition 4.1, $\hat{\varphi}_{IVA}$ is consistent and $\Psi_\Omega (\hat{\varphi}_{IVA} - \varphi) \xrightarrow{d} N\left(0; P_\Omega^{-1} \Gamma_\Omega P_\Omega^{-1}\right)$ as $N, T \to \infty$, where*

$$\begin{aligned} \Gamma_\Omega &= \lim_{N,T\to\infty} \left[(NT)^{-2} \Psi_\Omega R' \mathcal{P}_V \Omega_\eta^{-1} \mathcal{P}_V R \Psi_\Omega\right], \\ \Psi_\Omega &= Diag_{j=i,p,r} \left[\sqrt{M_j} I_{\dim R_{\Omega j}}\right] \\ R &= (R_{\Omega i}, R_{\Omega p}, R_{\Omega r}). \end{aligned}$$

Two cases can be distinguished:

1. *If $Z = \iota_{NT}$, then $M_i = N$, $M_p = T$ and $M_r = NT$,*

$$P_\Omega^{-1} \Gamma_\Omega P_\Omega^{-1} = Diag\left[\sigma_\lambda^2 P_i^{-1}, \sigma_\tau^2 P_p^{-1}, \sigma_\xi^2 P_r^{-1}\right]$$

and

(a) if $N = O(T)$*, we have that* $R_{\Omega i} = \left(R^{a\ddot{b}\ddot{c}}, R^{\dot{a}\ddot{b}\ddot{c}}\right)$*,* $R_{\Omega p} = \left(R^{b\ddot{c}}, R^{\dot{b}\ddot{c}}\right)$ *and* $R_{\Omega r} = \left(R^{c}, R^{\dot{c}}\right)$*, or*

(b) if $T = o(N)$*, it holds that* $R_{\Omega i} = \left(R^{a\ddot{c}}, R^{\dot{a}\ddot{c}}\right)$*,* $R_{\Omega p} = \left(R^{\ddot{a}\ddot{b}\ddot{c}}, R^{\ddot{a}\dot{b}\ddot{c}}\right)$ *and* $R_{\Omega r} = \left(R^{c}, R^{\dot{c}}\right)$*;*

2. Otherwise (i.e. if $Z \neq \iota_{NT}$*), then* $M_i = M_p = M_r = NT$*.*

Remark 3. Case 1.a encompasses $N \sim T$, in which case the distinction between $R_{\Omega i}$ and $R_{\Omega p}$ is unnecessary, since the consistency rates M_i and M_p are identical. For consistency with their definition, the distinct notation of these subsets is maintained and the subsets R^{ab}, $R^{\dot{a}b}$, $R^{a\dot{b}}$ and $R^{\dot{a}\dot{b}}$ of R were arbitrarily allocated to $R_{\Omega i}$, which makes it identical to case 1.a.

Case one only occurs in a model where the only random coefficient is associated with the constant term. It is consistent with the result of Wyhowski (1994). Remark that for the error component model it holds that $\text{AVar}\left[\hat{\varphi}_{IV} - \varphi\right] \geq \text{AVar}\left[\hat{\varphi}_{IVA} - \varphi\right]$. Indeed, although the convergence rates of $\hat{\varphi}_{IV}$ and $\hat{\varphi}_{IVA}$ are identical for individual-specific, period-specific and 'pure' idiosyncratic variables, variables with two or more components belong to the group for which the convergence rate is *lowest* in the case of $\hat{\varphi}_{IV}$ (Theorem 4.1), whereas they belong to the group with the *highest* convergence rate in the case of $\hat{\varphi}_{IVA}$ (Theorem 4.2). Case two is the general case (Hsiao (1975)) in which we have that $\text{AVar}\left[\hat{\varphi}_{IV} - \varphi\right] > \text{AVar}\left[\hat{\varphi}_{IVA} - \varphi\right]$, since $\hat{\varphi}_{IVA}$ (case 2) always has a higher convergence rate than $\hat{\varphi}_{IV}$ (cases 2-4).

With the use of $\hat{\varphi}_{IV}$, we will now proceed to estimate $\Omega_\eta = \text{E}\left[HH'\right]$ (in the next subsection) in order to construct a feasible Aitken estimator (see subsection 4.4.4).

4.4.3 Covariance Estimation

The composite error term of (4.9) is given by $H = Z^{(D)}U^{(S)} - \sigma_{xu}I_{NT}$. I make the following simplifying assumption regarding its covariance matrix.

Assumption E. The covariance matrix of the error term H of equation (4.9), obeys the following restriction

$$\text{Var}\left[H\,|V\right] \;=\; \text{E}\left[V^{(D)}\left(I_N \otimes J_T \otimes \tilde{\Sigma}_\lambda + J_N \otimes I_T \otimes \tilde{\Sigma}_\tau + I_{NT} \otimes \tilde{\Sigma}_\xi\right)V^{(D)\prime}\right].$$

Defining $\hat{Z} = \mathcal{P}_V Z$ and $\tilde{Z} = Z - \hat{Z}$, Assumptions A-D imply that

$$\begin{aligned}\mathrm{Var}\left[H\,|V\right] &= \hat{Z}^{(D)}\left(I_N \otimes J_T \otimes \Sigma_\lambda + J_N \otimes I_T \otimes \Sigma_\tau + I_{NT} \otimes \Sigma_\xi\right)\hat{Z}^{(D)\prime} \\ &\quad + \hat{Z}^{(D)}\mathrm{E}\left[U^{(S)}U^{(S)\prime}\tilde{Z}^{(D)\prime}\,|V\right] + \mathrm{E}\left[\tilde{Z}^{(D)}U^{(S)}U^{(S)\prime}\,|V\right]\hat{Z}^{(D)\prime} \\ &\quad + \mathrm{E}\left[\tilde{Z}^{(D)}U^{(S)}U^{(S)\prime}\tilde{Z}^{(D)\prime}\,|V\right] - \sigma_{xu}^2 I_{NT}. \end{aligned} \tag{4.15}$$

Assumption E now imposes the additional condition that the third and fourth order conditional moments appearing in (4.15) reduce to polynomials of maximal degree two in the instruments. As a consequence it holds that $\mathrm{E}\left[HH' \mid V\right]$ has the same structure as $\mathrm{E}\left[HH' \mid Z\right]$ would have in the standard random coefficient model, under the assumption of nonstochasticity of the regressors Z (as in Hsiao (1975)). The matrices of parameters that determine the variance of the disturbances are now given by the $k_V \times k_V$ matrices of $k_\eta = k_V\left(k_V+1\right)/2$ unknowns $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$, where the tilde reflects the facts that, first

$$\hat{Z}^{(D)}\left(I_N \otimes J_T \otimes \Sigma_\lambda + J_N \otimes I_T \otimes \Sigma_\tau + I_{NT} \otimes \Sigma_\xi\right)\hat{Z}^{(D)\prime}$$

is reparametrized in terms of V, and second that they have absorbed the influence of the last three terms that appear in (4.15). I propose the folllowing procedure that only makes use of the residuals $\hat{H} = Y - R\hat{\varphi}_{IV}$ to estimate $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$ consistently.

The matrix (4.15) contains $NT\left(N+T-1\right)$ nonzero elements, which lead to the following set of equations. First, the NT diagonal elements are stacked as[5]

$$H^{(\xi)} = V^{(\xi)}\left(\tilde{\Sigma}^{(\lambda)} + \tilde{\Sigma}^{(\tau)} + \tilde{\Sigma}^{(\xi)}\right) + E^{(\xi)}, \tag{4.16}$$

where $H^{(\xi)} = \left(H_1^{(\xi)\prime}, H_2^{(\xi)\prime}, \ldots, H_N^{(\xi)\prime}\right)'$, $V^{(\xi)} = \left(v_{11}^{(\xi)}, v_{12}^{(\xi)}, \ldots, v_{NT}^{(\xi)}\right)'$ and $E^{(\xi)} = H^{(\xi)} - \mathrm{E}\left[H^{(\xi)}\right]$ with $H_j^{(\xi)} = \hat{H}_j \odot \hat{H}_j = \left(\hat{\eta}_{j1}^2, \hat{\eta}_{j2}^2, \ldots, \hat{\eta}_{jT}^2\right)'$, $v_{it}^{(\xi)} = \underline{\mathrm{Vech}}\left[v_{it}v_{it}'\right]$ and $\tilde{\Sigma}^{(\xi)} = \mathrm{Vech}\left[\tilde{\Sigma}_\xi\right]$.

Next, the $O_1 = NT\left(T-1\right)/2$ elements representing correlations between different observations on the same individual are stacked as

$$H^{(\lambda)} = V^{(\lambda)}\tilde{\Sigma}^{(\lambda)} + E^{(\lambda)}, \tag{4.17}$$

where $H^{(\lambda)} = \left(H_1^{(\lambda)}, H_2^{(\lambda)}, \ldots, H_N^{(\lambda)}\right)$, $V^{(\lambda)} = \left(V_1^{(\lambda)}, V_2^{(\lambda)}, \ldots, V_N^{(\lambda)}\right)$ and $E^{(\lambda)} = H^{(\lambda)} -$

[5] The matrix operation $\underline{\mathrm{Vech}}\left[X\right]$ is defined for a $J \times J$ matrix X as $\mathrm{Vech}\left[X + X' - I_J \odot X\right]$, where $\odot$ denotes the Hadamard product,

$\mathrm{E}\left[H^{(\lambda)}\right]$ with $H_j^{(\lambda)} = (\hat{\eta}_{j1}\hat{\eta}_{j2}, \hat{\eta}_{j1}\hat{\eta}_{j3}, \ldots, \hat{\eta}_{j,T-1}\hat{\eta}_{jT})'$, $V_j^{(\lambda)} = \left(v_{j;12}^{(\lambda)}, v_{j;13}^{(\lambda)}, \ldots, v_{j;T-1,T}^{(\lambda)}\right)'$ $v_{j;st}^{(\lambda)} = \underline{\mathrm{Vech}}\left[v_{js}v_{jt}'\right]$ and $\tilde{\Sigma}^{(\lambda)} = \mathrm{Vech}\left[\tilde{\Sigma}_\lambda\right]$.

Finally, the $O_2 = N(N-1)T/2$ elements representing correlations between different individuals observed at the same moment are stacked as

$$H^{(\tau)} = V^{(\tau)}\tilde{\Sigma}^{(\tau)} + E^{(\tau)}, \tag{4.18}$$

where $H^{(\tau)} = \left(H_1^{(\tau)}, H_2^{(\tau)}, \ldots, H_T^{(\tau)}\right)$, $V^{(\tau)} = \left(V_1^{(\tau)}, V_2^{(\tau)}, \ldots, V_T^{(\tau)}\right)$ and $E^{(\tau)} = H^{(\tau)} - \mathrm{E}\left[H^{(\tau)}\right]$ with $H_s^{(\tau)} = (\hat{\eta}_{1s}\hat{\eta}_{2s}, \hat{\eta}_{1s}\hat{\eta}_{3s}, \ldots, \hat{\eta}_{N-1,s}\hat{\eta}_{Ns})'$, $V_s^{(\tau)} = \left(v_{s;12}^{(\tau)}, v_{s;13}^{(\tau)}, \ldots, v_{s;N-1,N}^{(\tau)}\right)'$ $v_{s;ij}^{(\tau)} = \underline{\mathrm{Vech}}\left[v_{is}v_{js}'\right]$ and $\tilde{\Sigma}^{(\tau)} = \mathrm{Vech}\left[\tilde{\Sigma}_\tau\right]$. All this can be summarized as

$$H^{(T)} = V^{(T)}\tilde{\Sigma}^{(T)} + E^{(T)},$$

where $H^{(T)} = \left(H^{(\lambda)\prime}, H^{(\tau)\prime}, H^{(\xi)\prime}\right)'$, $\tilde{\Sigma}^{(T)} = \left(\tilde{\Sigma}^{(\lambda)\prime}, \tilde{\Sigma}^{(\tau)\prime}, \tilde{\Sigma}^{(\xi)\prime}\right)'$, $E^{(T)} = \left(E^{(\lambda)\prime}, E^{(\tau)\prime}, E^{(\xi)\prime}\right)'$ and

$$V^{(T)} = \begin{pmatrix} V^{(\lambda)} & \mathbf{O}_{O_1\times O_3} & \mathbf{O}_{O_1\times O_3} \\ \mathbf{O}_{O_2\times O_3} & V^{(\tau)} & \mathbf{O}_{O_2\times O_3} \\ V^{(\xi)} & V^{(\xi)} & V^{(\xi)} \end{pmatrix},$$

where $\mathbf{O}_{J\times K}$ is equal to a matrix of zeroes. The unknown parameters of (4.15) can now be estimated by

$$\begin{aligned} \hat{\tilde{\Sigma}}^{(T)} &= \left(V^{(T)\prime}V^{(T)}\right)^{-1}V^{(T)\prime}H^{(T)} \\ &= \tilde{\Sigma}^{(T)} + \left(V^{(T)\prime}V^{(T)}\right)^{-1}V^{(T)\prime}E^{(T)}. \end{aligned} \tag{4.19}$$

The proposed estimator is biased, but consistent with convergence rate given in the following proposition.

Proposition 4.2. *The order of convergence for the proposed estimator of the variance components, $\hat{\tilde{\Sigma}}^{(T)}$, is given by*

$$\begin{aligned} \sqrt{\min(N,T^2)}\left(\hat{\tilde{\Sigma}}^{(\lambda)} - \tilde{\Sigma}^{(\lambda)}\right) &= O_p(1), \\ \sqrt{\min(N^2,T)}\left(\hat{\tilde{\Sigma}}^{(\tau)} - \tilde{\Sigma}^{(\tau)}\right) &= O_p(1), \\ \sqrt{\min(N,T)}\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right) &= O_p(1). \end{aligned}$$

The proof is given in Appendix C.

Rearranging the elements of $\hat{\tilde{\Sigma}}^{(T)}$, we obtain consistent estimates of $\tilde{\Sigma}_\lambda$, $\tilde{\Sigma}_\tau$ and $\tilde{\Sigma}_\xi$, allowing the construction of

$$\hat{\Omega}_\eta \;=\; \mathrm{E}\left[V^{(D)}\left(I_N \otimes J_T \otimes \hat{\tilde{\Sigma}}_\lambda + J_N \otimes I_T \otimes \hat{\tilde{\Sigma}}_\tau + I_{NT} \otimes \hat{\tilde{\Sigma}}_\xi\right) V^{(D)\prime}\right].$$

In Appendix C an expression is given which simplifies the inversion of the $NT \times NT$ matrix $\hat{\Omega}_\eta$ to the inversion of the NT elements of a diagonal matrix, the inversion of a $NK \times NK$ matrix and the inversion of a $TK \times TK$ matrix, using a slight modification of an expression given by Hsiao (1975). Note that the proposed method to estimate the covariance matrix requires only two regressions, a first to estimate φ, and the second to estimate $\tilde{\Sigma}^{(T)}$.

4.4.4 Feasible Aitken Estimation

Using $\hat{\Omega}_\eta$ from the previous paragraph, or any other consistent estimator of φ, the feasible IV version of the Aitken estimator (4.14) is given by

$$\hat{\varphi}_{FIVA} \;=\; \left(R'\mathcal{P}_V\hat{\Omega}_\eta^{-1}R\right)^{-1} R'\mathcal{P}_V\hat{\Omega}_\eta^{-1}Y, \tag{4.20}$$

for which the asymptotic properties are given in the following theorem.

Theorem 4.3. *Under Assumption A-E, under the sufficient identification conditions from Proposition 4.1 and using the covariance estimator from Proposition 4.2,* $\Psi_{\hat{\Omega}}\left(\hat{\varphi}_{FIVA} - \varphi\right) = O_p\left(1\right)$ *as* $N, T \to \infty$*, where*

$$\begin{aligned} \Psi_{\hat{\Omega}} &= Diag_{j=i,p,r}\left[\sqrt{M_j} I_{\dim R_j}\right] \\ R &= \left(R_{\hat{\Omega}i}, R_{\hat{\Omega}p}, R_{\hat{\Omega}r}\right). \end{aligned}$$

Furthermore, two cases are discernable:

1. *If* $Z = \iota_{NT}$*, then* $M_i = N$*,* $M_p = T$ *and* $M_r = NT$*. Furthermore, it holds that*

$$\Psi_{\hat{\Omega}}\left(\hat{\varphi}_{FIVA} - \varphi\right) \xrightarrow{d} N\left(0; P_\Omega^{-1}\Gamma_\Omega P_\Omega^{-1}\right),$$

with $P_\Omega^{-1}\Gamma_\Omega P_\Omega^{-1} = Diag\left[\sigma_\lambda^2 P_i^{-1}, \sigma_\tau^2 P_p^{-1}, \sigma_\xi^2 P_r^{-1}\right]$*. In addition*

(a) *if* $N = O(T)$, *then we have that* $R_{\hat{\Omega}i} = \left(R^{a\ddot{b}\ddot{c}}, R^{\dot{a}\ddot{b}\ddot{c}}\right)$, $R_{\hat{\Omega}p} = \left(R^{b\ddot{c}}, R^{\dot{b}\ddot{c}}\right)$ *and* $R_{\hat{\Omega}r} = \left(R^{c}, R^{\dot{c}}\right)$,

(b) *if* $T = o(N)$, *then it holds that* $R_{\hat{\Omega}i} = \left(R^{a\ddot{c}}, R^{\dot{a}\ddot{c}}\right)$, $R_{\hat{\Omega}p} = \left(R^{\ddot{a}b\ddot{c}}, R^{\ddot{a}\dot{b}\ddot{c}}\right)$ *and* $R_{\hat{\Omega}r} = \left(R^{c}, R^{\dot{c}}\right)$;

2. *Otherwise (i.e. if* $Z \neq \iota_{NT}$*),* $N_i = N_p = N_r = \min\left(N^2, T^2\right)$.

Remark 4. Case 1.a encompasses $N \sim T$, in which case the distinction between $R_{\Omega i}$ and $R_{\Omega p}$ is unnecessary, since the consistency rates M_i and M_p are identical. For consistency with their definition, the distinct notation of these subsets is maintained and the subsets R^{ab}, $R^{\dot{a}b}$, $R^{a\dot{b}}$ and $R^{\dot{a}\dot{b}}$ of R were arbitrarily allocated to $R_{\Omega i}$, which makes it identical to case 1.a.

The proof of Theorem 4.3 is given in Appendix C. Part one pertains to the error component model and it states that the asymptotic distribution of the feasible estimator is identical to that of the estimator with known covariance matrix. Part two establishes the convergence rate for the feasible IV Aitken estimator that uses the covariance estimator from subsection 4.4.3. The latter consistency rate is not identical to the consistency rate of the IV Aitken estimator with known covariance matrix (4.14), except in the special case when $N \sim T$, since then it holds that $NT = \min\left(N^2, T^2\right)$. The latter result is consistent with Hsiao (1976).

The importance of Theorem 4.3 lies first of all, in the fact that it provides convergence rates for the feasible two-stage Aitken estimator without imposing a restriction on the ratio N/T. A second important finding is that the feasible Aitken estimator does not converge to the Aitken estimator when we consider a (correlated) random coefficient model where the random parts of the coefficients follow a two-way error specification.

In accordance with the results of Swamy (1970), Kelejian and Stephan (1983) noted that the convergence rate reported by Hsiao (1975) is counter-intuitive. As Mandy and Martins-Filho (1994) remark, the asymptotic equivalence between Aitken and feasible Aitken estimators is often proved by relying on Theorem 3 of Fuller and Battese (1973), which the former authors disprove by means of a counter-example. More specifically, the conditions under which Fuller and Battese state their theorem to be valid are much too general. The asymptotic equivalence between Aitken and feasible Aitken estimators in the case of a diagonal disturbance covariance matrix (see Carroll and Ruppert (1982) and Crocket (1985)) cannot be extended to the case of a non-diagonal covariance matrix without imposing restrictions on the off-diagonal elements.

Sufficient conditions for the asymptotic equivalence of $\hat{\varphi}_{IVA}$ and $\hat{\varphi}_{FIVA}$ are

$$\operatorname*{plim}_{N,T\to\infty}\left[(NT)^{-1}R'\mathcal{P}_V\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)R\right] = 0 \tag{4.21}$$

$$\operatorname*{plim}_{N,T\to\infty}\left[(NT)^{-1}\Psi_\Omega R'\mathcal{P}_V\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)H\right] = 0 \tag{4.22}$$

(White (1984)). Mandy and Martins-Filho (1994) list conditions under which consistency of the covariance estimator implies both conditions (4.21) and (4.22) and thus asymptotic equivalence between $\hat{\varphi}_{IVA}$ and $\hat{\varphi}_{FIVA}$. Unfortunately, one of their conditions requires the number of nonzero elements in each column (and thus row) of $\hat{\Omega}_\eta^{-1}$ to be uniformly bounded as $N,T\to\infty$. Consequently, their Theorem 1 cannot be used to prove asymptotic equivalence of $\hat{\varphi}_{IVA}$ and $\hat{\varphi}_{FIVA}$ here. For the special case of the error components model (part 1 in Theorem 4.3), (4.21) and (4.22) are easily proved. For a general CRC model, (4.21) also holds, but (4.22) is violated and the correct order of convergence is shown to be $\min(N^2,T^2)$.

4.5 Concluding Remarks

In this chapter some estimators are presented for a linear CRC model with panel data, where each random slope can be written as a linear function of some variables and an additive error term that consists of three components, all three of which can be correlated with some of the regressors. I propose to estimate the ATE by standard IV methods, using as instruments cross-products of the set of HT instruments of the model variables and the set of HT instruments of the slope variables, after removal of product variables that occur more than once. Sufficient identification conditions are obtained for this estimator, which is shown to be asymptotically normal. Also the properties of the Aitken and feasible Aitken IV estimators were investigated. In the special case of the error-component model they were shown to converge to the same asymptotic distribution. In the case of a general CRC model, however, they were shown to be asymptoticaly nonequivalent and the rate of convergence of the feasible Aitken IV estimator was obtained.

Chapter 5

Unit Root Tests in Panel Data Models with Serially Correlated Errors and Fixed T

Abstract. This paper derives unit root tests for $AR(1)$ panel data models with $AR(1)$ errors and fixed time dimension. The limiting distributions of the test statistics (for increasing N and fixed T) are shown to be normal. Closed-form expressions for the first and second moments of the test statistics are derived. Heterogeneous initial conditions and drift in the data generating process are taken into account by including fixed effects and individual-specific linear time trends in the regression. This inclusion makes the least squares estimators of the autoregressive parameters inconsistent for fixed T, and appropriate bias-corrections are proposed.
JEL classification: C22, C23
Keywords: Unit root; Panel Data; Serial correlation

5.1 Introduction

A considerable amount of econometric research is devoted to the analysis of integrated time series. In the univariate case, the augmented Dickey-Fuller (1979) (ADF) and the semi-parametric Phillips-Perron (1988) test procedures are well-known tests for the presence of a unit root. Unfortunately, in finite samples these tests have limited power against near unit root alternatives. The extension of these techniques to panel data generates more powerful

unit root tests, when the data are independently distributed over individuals. Recently, unit root test procedures for dynamic panel data models have been suggested (for recent overviews see Banerjee (1999), Baltagi and Kao (2001)). These procedures are based on test statistics that converge to a normal distribution in this context.

Quah (1994) uses random field methods to analyze a model with IID disturbances, under the assumption that both N and T tend to infinity at a fixed rate. Unfortunately, this approach is not suitable for the analysis of more general model specifications that allow, for example, for individual-specific fixed effects. Levin and Lin (1993) consider a data generating process (DGP) that includes individual drift parameters and individual deterministic trends. Furthermore, they allow the errors to be heterogeneously serially correlated. Assuming that both N and T go to infinity while $N/T \to 0$, they show that the inclusion of individual fixed effects and/or time trends in the regression causes the least squares estimator of the autoregressive parameter to be downwardly biased. Considering the same class of models, Im et al. (2003) propose a test based on the average of N individual ADF statistics. This allows them to weaken the assumption of Levin and Lin to $\sqrt{N}/T \to 0$. They also show that their tests are valid for fixed T, provided that the error terms are serially uncorrelated and independent across individuals.

All above-mentioned unit root test statistics for panel data models have limiting normal distributions, whose moments can be derived for both $N, T \to \infty$. In such multi-indexed asymptotics several approaches are possible, depending on the passage to infinity of both indexes (Phillips and Moon (1999a,b)).

A completely different approach builds on results obtained in the context of 'Meta Analysis', where the main idea rests on the combination of observed p-values of independent tests to assess an overall hypothesis. The Fisher (1932) test is a non-parametric, exact test that is distributed as χ^2_{2N} (Maddala and Wu (1999), Choi (2001)). This class of tests allows for finite N, but assumes $T \to \infty$. For small values of T, the authors propose using simulated moments when implementing their tests.

In contrast with the previous approaches, Harris and Tzavalis (1999) develop asymptotic unit root tests for first-order autoregressive panel data models with serially uncorrelated errors, under the assumption that $N \to \infty$, while T is fixed and identical for all individuals. They show that the limiting distribution of these test statistics is standard normal, and that the convergence rate is the same as for stationary panel data, i.e. $\sqrt{N}$. Their test statistics are based on the standardized least squares estimators of the autoregressive coefficients, and their distribution under the null hypothesis is invariant with respect to the nuisance

parameters of the DGP. These nuisance parameters are the variances of the error terms, the starting values and the individual drift parameters. When fixed effects or individual deterministic trends are included in the regression, the LS estimators are asymptotically downwardly biased, even as $T \to \infty$, which contrasts with the stationary case. Harris and Tzavalis (1998) extend the approach with finite T to the case of heteroskedastic $MA(1)$ errors. This extension, however, is only derived in the context of a DGP without drift. Kruiniger and Tzavalis (2001) allow the disturbances to follow an $MA(T-1)$ process and the DGP to exhibit drift.

In this chapter I adapt the approach of Harris and Tzavalis (1999) by allowing the error terms to be generated by an $AR(1)$ process. The extension of uncorrelated errors to $AR(1)$ disturbances in a panel data context corresponds to the extension of the DF-test to the $ADF(1)$-test in a time series context, keeping in mind that the panel data test statistics converge in distribution to the normal distribution. The DGP in our model is also allowed to include non-zero starting values and individual-specific drift terms. It is found that neglecting positive autocorrelation in the errors reduces the size of the test, while negative autocorrelation inflates the size of the test. Given the fact that in most empirical studies N is larger than T, tests for fixed T seem particularly useful.

Below, I consider three different data generating processes, all of which have a unit root under the null hypothesis, and $AR(1)$ errors. The first DGP is an $AR(1)$ process where the starting values are assumed to be zero, and is estimated by the pooled least squares estimator. The second DGP has non-zero starting values and is estimated by the fixed effects estimator. The third DGP has non-zero starting values and a heterogeneous drift parameter and it is estimated by a least squares estimator that includes both fixed effects and an individual-specific linear trend. For the first DGP and the pooled LS estimator, analytical expressions for the bias (which is zero) and the variance of the unit root parameter (φ) estimator are derived. The variance depends on the autoregressive coefficient (ρ) of the error term, and can be consistently estimated either by substituting a consistent estimate for ρ, or by an estimator that does not rely on the analytical expression for the variance. This leads to two different unit root tests. For the second and third DGP the LS estimators for $\hat{\varphi}$ and $\hat{\rho}$ are shown to be inconsistent and hence need to be modified. A corrected estimate of φ is obtained by inserting a corrected estimate of ρ in the expression for the probability limit of $\hat{\varphi}$. Such an estimate of ρ follows upon inverting the probability limit of $\hat{\rho}$, which is shown to be monotonically increasing in ρ. A second modification involves adjusting the variance of the test statistic, to take into account the estimated nature

of the bias-correction that is used.

The organization of the chapter is as follows. Section 5.2 presents the three different models and the associated LS estimates. In section 5.3 a unit root test is derived for each of these models. Section 5.4 provides some concluding remarks. Proofs are given in Appendix D.

5.2 Theoretical Framework

Consider the following three data generating processes

$$\text{DGP 1} \quad : \quad y_{i,t} = \varphi y_{i,t-1} + u_{it}, \quad \text{with } y_{i0} = 0, \tag{5.1}$$

$$\text{DGP 2} \quad : \quad y_{i,t} = \varphi y_{i,t-1} + u_{it}, \tag{5.2}$$

$$\text{DGP 3} \quad : \quad y_{i,t} = \tau_i + \varphi y_{i,t-1} + u_{it}, \tag{5.3}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and the error term follows an $AR(1)$ process starting in some given period $S \leq 1$

$$u_{it} \quad = \quad \begin{cases} 0, & \text{if } t < s, \\ \rho u_{i,t-1} + \varepsilon_{it}, & \text{otherwise,} \end{cases} \tag{5.4}$$

with $\rho < 1$. Our main interest lies in testing the hypothesis of a unit root, corresponding to $\varphi = 1$. The difference with previous unit root tests for panels with fixed time dimension (Harris and Tzavalis (1998, 1999)), is the fact that I allow the error terms to follow a first-order autoregressive specification. The formulation of (5.4) encompasses two specific cases of interest, i.e. the error process starting in period 1 ($S = 1$) and the error process starting in the infinite past ($S = -\infty$).

In order to obtain a test statistic, a first step is to estimate the following regressions, respectively, by least squares

$$\text{Model 1} \quad : \quad y_{i,t} = \varphi y_{i,t-1} + \rho \Delta y_{i,t-1} + \varepsilon_{it}, \qquad t = 2, \ldots, T \tag{5.5}$$

$$\text{Model 2} \quad : \quad y_{i,t} = \alpha_i + \varphi y_{i,t-1} + \rho \Delta y_{i,t-1} + \varepsilon_{it}, \qquad t = 3, \ldots, T \tag{5.6}$$

$$\text{Model 3} \quad : \quad y_{i,t} = \alpha_i + \beta_i t + \varphi y_{i,t-1} + \rho \Delta y_{i,t-1} + \varepsilon_{it}, \; t = 3, \ldots, T \tag{5.7}$$

where $i = 1, \ldots, N$. It is assumed that the data are observed for $t = 1, \ldots T$. The assumption $y_{i0} = 0$ in (5.1) allows us to gain an additional observation at $t = 0$ for the

estimation of Model 1, compared to the estimation of Models 2 and 3. The asymptotic distributions of the OLS estimators $\hat{\varphi}$ and $\hat{\rho}$ will be derived for the Models 1 to 3, under the following assumption.

(A1) The ε_{it} are IID with $\mathrm{E}[\varepsilon_{it}] = 0$, $\mathrm{E}[\varepsilon_{it}^2] = \sigma_\varepsilon^2 > 0$ and $\mathrm{E}[\varepsilon_{it}^4] < \infty$,

(A2) the individual-specific vectors $\begin{pmatrix} y_{i0}, & \tau_i \end{pmatrix}'$ are IID, independently of ε_{it},

(A3) $\mathrm{Var}\left[\begin{pmatrix} y_{i0} & \tau_i \end{pmatrix}'\right] = \begin{pmatrix} \sigma_{y_0}^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}$ with $\mathrm{E}[y_{i0}^4] < \infty$ and $\mathrm{E}[\tau_i^4] < \infty$.

Assumptions (A1)-(A3) imply that the error terms ε_{it} are independent across all observations, and that all observables are independent across all individual units, i. These assumptions allow the application of the Law of Large Numbers and the Lindeberg-Lévy Central Limit Theorem for $N \to \infty$. Assumption (A1) further implies that the tests are comparable with the augmented Dickey-Fuller tests for a single time series where the autocorrelation of the error term is restricted to be of first order. Assumption (A2) implies that (y_{i0}, τ_i) is uncorrelated with ε_{it}, and (A3) implies that y_{i0} and τ_i are uncorrelated, but both assumptions allow for arbitrary correlation between the unobserved effects (y_{i0} and τ_i) and the observed explanatory variables ($y_{i,t-1}$ and $\Delta y_{i,t-1}$) (Wooldridge (2002), p. 252). Remark that Assumptions (A2)-(A3) do not exclude the special case that y_{i0} and τ_i are fixed, as is assumed by Harris and Tzavalis (1999).

The unit root hypothesis is given by

$$H_0 : \varphi = 1 \quad \text{versus} \quad H_A : \varphi < 1.$$

The three DGP's (5.1)-(5.3) are the panel data counterparts of the single time series DGP's considered by Phillips and Perron (1988) in the development of their unit root test. The unit root test considered here is based upon the standardized coefficient statistic, which is similar to other tests for finite T (Harris and Tzavalis (1998, 1999)).

Consider the least squares estimators $\hat{\varphi}$ and $\hat{\rho}$ of the regression models (5.5)-(5.7), under the assumption that the data are generated by the processes (5.1)-(5.3), respectively. Under H_0 we have, by the Frisch-Waugh (1933)-Lovell (1963) Theorem,

$$\begin{pmatrix} \hat{\varphi} \\ \hat{\rho} \end{pmatrix} = \begin{pmatrix} 1 \\ \rho \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^N y'_{i,-1} Q_{T-j} y_{i,-1} & \sum_{i=1}^N y'_{i,-1} Q_{T-j} \Delta y_{i,-1} \\ \sum_{i=1}^N y'_{i,-1} Q_{T-j} \Delta y_{i,-1} & \sum_{i=1}^N \Delta y'_{i,-1} Q_{T-j} \Delta y_{i,-1} \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{i=1}^N y'_{i,-1} Q_{T-j} \varepsilon_i \\ \sum_{i=1}^N \Delta y'_{i,-1} Q_{T-j} \varepsilon_i \end{pmatrix} \tag{5.8}$$

where $j = 1$ for Model 1 and $j = 2$ for Models 2 and 3,

$$\begin{aligned} \varepsilon_i &= (\varepsilon_{i,j+1}, \ldots, \varepsilon_{i,T})', \quad y_{i,-1} = (y_{i,j}, \ldots, y_{i,T-1})', \\ \Delta y_{i,-1} &= (y_{i,j} - y_{i,j-1}, \ldots, y_{i,T-1} - y_{i,T-2})', \end{aligned} \tag{5.9}$$

and Q_T is a $T \times T$ idempotent matrix, for the respective models given by

$$\text{Model 1} \ : \ Q_T = I_T, \tag{5.10}$$

$$\text{Model 2} \ : \ Q_T = I_T - \frac{\iota_T \iota_T'}{T}, \tag{5.11}$$

$$\text{Model 3} \ : \ Q_T = I_T - Z_T \left(Z_T' Z_T\right)^{-1} Z_T', \tag{5.12}$$

where ι_T is a $T \times 1$ vector of ones, $Z_T = (\iota_T, \kappa_T)$ and $\kappa_T' = (1, 2, \ldots, T)$. For Model 1, estimator (5.8) is the least squares (LS) pooled estimator $(\hat{\varphi}, \hat{\rho})'_{LSP}$; for Model 2 it is the LS dummy variable estimator $(\hat{\varphi}, \hat{\rho})'_{LSDV}$; and for Model 3 it is the LS dummy variable with trend estimator $(\hat{\varphi}, \hat{\rho})'_{LSDT}$. Letting

$$\begin{array}{lllll} p_i &=& \sigma_\varepsilon^{-2} y_{i,-1}' Q_{T-j} y_{i,-1}, & \bar{p} &= \frac{1}{N}\sum_{i=1}^N p_i, \\ q_i &=& \sigma_\varepsilon^{-2} \Delta y_{i,-1}' Q_{T-j} y_{i,-1}, & \bar{q} &= \frac{1}{N}\sum_{i=1}^N q_i, \\ r_i &=& \sigma_\varepsilon^{-2} \Delta y_{i,-1}' Q_{T-j} \Delta y_{i,-1}, & \bar{r} &= \frac{1}{N}\sum_{i=1}^N r_i, \\ v_i &=& \sigma_\varepsilon^{-2} y_{i,-1}' Q_{T-j} \varepsilon_i, & \bar{v} &= \bar{v} = \frac{1}{N}\sum_{i=1}^N v_i, \\ w_i &=& \sigma_\varepsilon^{-2} \Delta y_{i,-1}' Q_{T-j} \varepsilon_i, & \bar{w} &= \frac{1}{N}\sum_{i=1}^N w_i, \end{array} \tag{5.13}$$

(5.8) can be rewritten as

$$\begin{pmatrix} \hat{\varphi} \\ \hat{\rho} \end{pmatrix} = \begin{pmatrix} 1 \\ \rho \end{pmatrix} + \begin{pmatrix} \bar{p} & \bar{q} \\ \bar{q} & \bar{r} \end{pmatrix}^{-1} \begin{pmatrix} \bar{v} \\ \bar{w} \end{pmatrix}, \tag{5.14}$$

from which the limiting behavior of $\hat{\varphi}$ and $\hat{\rho}$ has to be determined. This will involve calculating the probability limits

$$(P, Q, R, V, W) = \operatorname{plim} (\bar{p}, \bar{q}, \bar{r}, \bar{v}, \bar{w}), \tag{5.15}$$

and the bias

$$(b_\varphi, b_\rho) = \operatorname{plim} (\hat{\varphi} - 1, \hat{\rho} - \rho).$$

Throughout the chapter, the asymptotics refer to limits as $N \to \infty$ while T is kept fixed,

unless otherwise stated.

5.3 Test Statistics

In this section the unit root tests for the different models will be presented. For Model 1, the bias and variance of the 'unit root'-parameter estimator, $\hat{\varphi}$, are derived, and the test statistic is the standardized parameter estimate. In addition, it is shown that the t-test, based on standard regression output, gives correct asymptotic inference for H_0. For Models 2&3 a somewhat different approach is followed, because a bias-correction is needed. For Model 1 a comparison is made with a test statistic based on a regression that does not take into account the serial correlation of the error terms (Harris and Tzavalis (1999)).

The limiting distributions of the estimators depend on S, the starting period of the $AR(1)$ error process, and in particular on the quantity $\omega = \omega(S) = \sum_{s=S}^{0} \rho^{2-2s}$ (with the convention that $\sum_{i=a}^{b}(\cdot) = 0$ when $a > b$). Note that

$$0 = \omega(1) \leq \omega(0) \leq \ldots \leq \omega(-\infty) = \frac{\rho^2}{1-\rho^2}$$

and that $\rho = 0$ implies $\omega = 0$.

5.3.1 Model 1

Model 1 consists of the true process given by (5.1), and is estimated by the pooled regression in (5.5). In the following theorem the limiting distribution of the pooled LS estimator is given.

Theorem 5.1. *If Assumption (A1) holds and $\varphi = 1$, then*

$$\sqrt{N}\begin{pmatrix} \hat{\varphi} - 1 \\ \hat{\rho} - \rho \end{pmatrix}_{LSP} \xrightarrow{d} N(0, \Sigma)$$

with

$$\Sigma = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1}$$

and

$$\begin{aligned} P &= \sum_{t=1}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2 + \omega\sum_{t=1}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2 \\ Q &= \sum_{t=1}^{T-1}\sum_{s=1}^{t}\rho^{s-1}\frac{1-\rho^s}{1-\rho} + \omega\sum_{t=1}^{T-1}\rho^{t-1}\frac{1-\rho^t}{1-\rho} \\ R &= \sum_{t=1}^{T-1}\sum_{s=1}^{t}\rho^{2s-2} + \omega\sum_{t=1}^{T-1}\rho^{2t-2}. \end{aligned}$$

Remark. The summations appearing in P, Q and R can be eliminated. The resulting expressions are given in Appendix D.

In view of Theorem 5.1, I propose the unit root test statistic

$$T_1 = (\hat{\varphi}_{LSP} - 1)\sqrt{\frac{N\left(\hat{P}\hat{R} - \hat{Q}^2\right)}{\hat{R}}},$$

where $\hat{P}$, $\hat{Q}$ and $\hat{R}$ are P, Q and R evaluated at $\hat{\rho}_{LSP}$. In contrast with a regression model with serially uncorrelated errors (Harris and Tzavalis (1999)), the nuisance parameter ρ appears in σ^2_{φ}, the limiting variance of $\sqrt{N}\hat{\varphi}_{LSP}$. However, since $\hat{\rho}_{LSP}$ is a $\sqrt{N}$-consistent estimator for ρ under H_0, σ^2_{φ} is estimated $\sqrt{N}$-consistently by substituting $\hat{\rho}_{LSP}$ for ρ. Hence, the resulting statistic, T_1, is asymptotically distributed as $N(0,1)$.

If we set ρ equal to zero, Σ reduces to

$$\Sigma = \begin{pmatrix} \frac{T}{2}(T-1) & (T-1) \\ (T-1) & (T-1) \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2}{(T-1)(T-2)} & -\frac{2}{(T-1)(T-2)} \\ -\frac{2}{(T-1)(T-2)} & \frac{T}{(T-1)(T-2)} \end{pmatrix}.$$

Taking the limit for $T \to \infty$ we have, by Theorem 5.1,

$$\sqrt{N}\begin{pmatrix} T(\hat{\varphi}-1) \\ \sqrt{T}(\hat{\rho}-\rho) \end{pmatrix}_{LSP} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

The existence of the sequential limit and the nature of DGP 1 allow us to conclude that

$$\sqrt{N}\left(T(\hat{\varphi}-1), \sqrt{T}(\hat{\rho}-\rho)\right)'_{LSP}$$

converges for $(N,T) \to \infty$ to the stated normal distribution, irrespective of the expan-

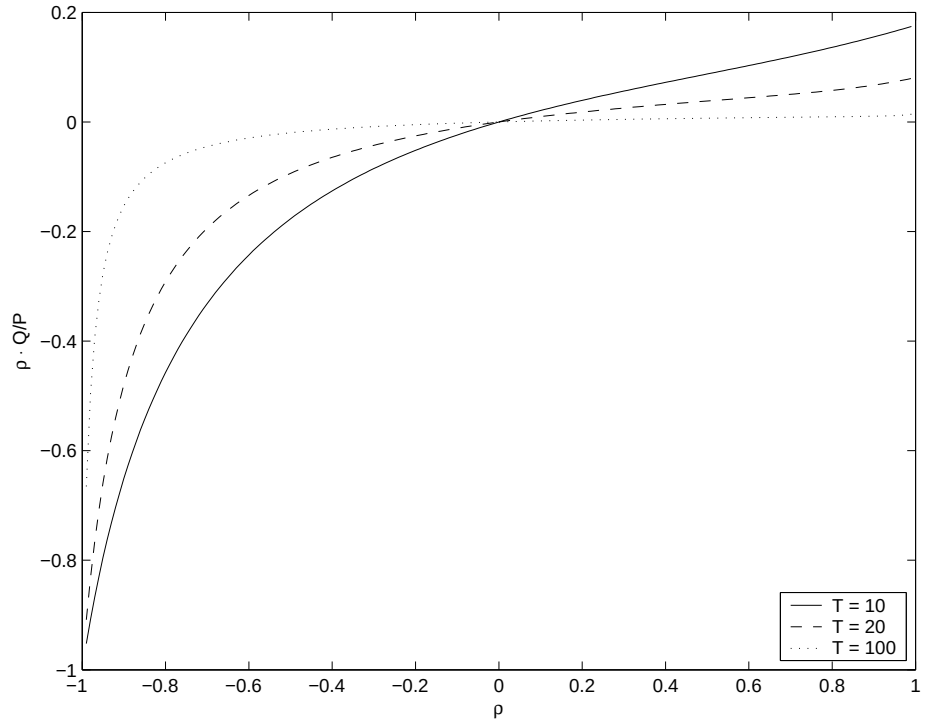

Figure 5.1: The Inconsistency of $\hat{\varphi}_{LSP;HT}$ due to Neglecting Serial Correlation.

sion rate between N and T (Phillips and Moon (1999a), Remark (a) on p. 1074). The convergence rates of $\hat{\rho}_{LSP}$ ($\sqrt{NT}$) and $\hat{\varphi}_{LSP}$ ($T\sqrt{N}$) are consistent with results reported previously by Levin and Lin (1993), Quah (1994) and Harris and Tzavalis (1999). The latter convergence rate is also known as 'super consistency'.

We will now discuss our test statistic when the error process starts at $S = 1$, and compare it with the test statistic proposed by Harris and Tzavalis (1999), both in the absence and presence of serial correlation. This discussion is based on the following lemma.

Lemma 5.1. *Let $Q = Q(\rho, T)$ and $P = P(\rho, T)$ be as in Theorem 5.1, and let $S = 1$. Then*

$$Q(\rho, T) \;>\; 0 \tag{5.16}$$

and

$$P(\rho, T) \gtreqless \frac{T(T-1)}{2} \iff \rho \gtreqless 0. \tag{5.17}$$

The pooled LS estimator of Harris and Tzavalis (1999) is

$$\hat{\varphi}_{LSP;HT} = \left(\sum_{i=1}^{N} y_{i,-1}^2 \right)^{-1} \left(\sum_{i=1}^{N} y_{i,-1} y_i \right)$$

and has variance $\sigma^2_{\varphi;HT} = 2T^{-1}(T-1)^{-1}$. If we compare the expression for σ^2_{φ} evaluated at $\rho = 0$ with $\sigma^2_{\varphi;HT}$, the ratio of both variances is given by

$$\frac{\sigma^2_{\varphi;HT}}{\sigma^2_{\varphi}\big|_{\rho=0}} = \frac{(T-2)}{T} < 1.$$

This means that in the absence of serial correlation $\hat{\varphi}_{LSP;HT}$ is a more efficient estimator for φ than $\hat{\varphi}_{LSP}$, but this efficiency gain disappears for large T. Note that the efficiency loss of $\hat{\varphi}_{LSP}$ is due to the fact $\Delta y_{i,t-1}$ is included in our regression model, so that an additional parameter has to be estimated and the effective number of observations along the time dimension is reduced by one.

The effect on the test statistic of neglecting serial correlation of the error term, i.e. estimating φ while restricting ρ to zero, is twofold. First it causes $\hat{\varphi}_{LSP;HT}$ to be an inconsistent estimator of φ. The magnitude of this inconsistency follows from Theorem 5.1 as

$$\operatorname*{plim}_{N,T\to\infty} \rho \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\Delta y_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}^2} = \operatorname*{plim}_{N,T\to\infty} \rho \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}^2} = \rho \frac{Q}{P}.$$

The first expression is the standard expression for the omitted variable bias when we regress $y_{i,t}$ on $y_{i,t-1}$ and omit $\Delta y_{i,t-1}$. The first equality follows from the fact that $y_{i0} = 0$ by assumption. The inconsistency is depicted in figure 5.1. For $\rho > 0$, it shifts the distribution of the test statistic with ρ restricted to zero to the right, meaning that the one-sided test of H_0 will be undersized (by Lemma 5.1). A second effect is that the reported variance, $\sigma^2_{\varphi;HT}$, inconsistently estimates the true variance of $\hat{\varphi}_{LSP;HT}$. The latter variance is given by

$$\operatorname{Var}\left[\hat{\varphi}_{LSP;HT}\right] = \left(\operatorname{E}\left[\sum_{t=1}^{T} y_{i,t-1}^2 \right] \right)^{-2} \operatorname{Var}\left[\sum_{t=1}^{T} y_{i,t-1}\varepsilon_{i,t} \right] = P^{-1}.$$

The effect of using $\sigma^2_{\varphi;HT}$ instead of $\operatorname{Var}\left[\hat{\varphi}_{LSP;HT}\right]$ is an overestimation of the true variance of $\hat{\varphi}_{LSP;HT}$ for $\rho > 0$, and vice versa (by Lemma 5.1). An overestimated variance will produce too thin tails *after standardization*, which results in an undersized test. The

combination of the inconsistency of $\hat{\varphi}_{LSP;HT}$ and the inconsistently estimated variance results unambiguously in an undersized unit root test when $\rho > 0$ and the test statistic of Harris and Tzavalis (1999) is used.

It needs to be pointed out that the asymptotic covariance matrix of $(\hat{\varphi}, \hat{\rho})'_{LSP}$, i.e. Σ, is consistently estimated by the usual covariance estimator, $(X'X)^{-1}\hat{\sigma}_{\varepsilon}^2$ in standard notation, although it is less efficient than the one I propose. As a practical consequence the standard t-test yields correct asymptotic inference for H_0.

In this subsection a test statistic was proposed for the simple model without individual-specific constant terms and with starting values equal to zero. Its main purpose was to make a comparison with the test statistic for a model with serially uncorrelated errors and fixed T (Harris and Tzavalis (1999)). The model considered is restrictive since it assumes that y_{i0} is zero for all i, which makes its empirical relevance limited. In the next section, a unit root test is derived that allows the DGP to have non-zero starting values, by including N fixed effects dummy variables in the regression equation.

5.3.2 Model 2

Model 2 is given by the DGP in (5.2) and is estimated by the LS dummy variables regression in (5.6). Note that, under H_0, $\alpha_i = 0$ in (5.6). The following theorem gives the limiting distribution of this estimator.

Theorem 5.2. *If Assumptions (A1)-(A3) hold and $\varphi = 1$, then*

$$\sqrt{N}\begin{pmatrix} \hat{\varphi} - 1 - b_{\varphi} \\ \hat{\rho} - \rho - b_{\rho} \end{pmatrix}_{LSDV} \xrightarrow{d} N(0, \Sigma)$$

where

$$\begin{pmatrix} b_{\varphi} \\ b_{\rho} \end{pmatrix} = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1} \begin{pmatrix} V \\ W \end{pmatrix} \tag{5.18}$$

with

$$\begin{aligned} V &= -\frac{1}{T-2}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}, \\ W &= -\frac{1}{T-2}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\rho^{t-s}, \end{aligned}$$

and

$$
\begin{aligned}
P &= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2 - \frac{2}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{(1-\rho^r)(1-\rho^{r+t-s})}{(1-\rho)^2} \\
&\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2 - \frac{\omega}{T-2}\left(\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\right)^2, \\
Q &= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\frac{\rho^{s-1}(1-\rho^s)}{1-\rho} - \frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho} \\
&\quad +\omega\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\rho^{t-1} - \frac{\omega}{T-2}\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}, \\
R &= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2} - \frac{2}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2} \\
&\quad +\omega\sum_{t=2}^{T-1}\rho^{2t-2} - \frac{\omega}{T-2}\left(\sum_{t=2}^{T-1}\rho^{t-1}\right)^2,
\end{aligned}
$$

and where

$$
\Sigma = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1} E[\xi_i\xi_i'] \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1},
$$

with

$$
\xi_i = \begin{pmatrix} v_i \\ w_i \end{pmatrix} - \begin{pmatrix} p_i & q_i \\ q_i & r_i \end{pmatrix}\begin{pmatrix} b_\varphi \\ b_\rho \end{pmatrix}, \tag{5.19}
$$

where p_i, q_i, r_i, v_i and w_i are defined in (5.13).

The bias of the *LSDV* estimator is known as the Nickell (1981) bias. Since Theorem 5.2 is only applicable under the condition that ρ is known, it needs to be adapted to obtain an operational test procedure. This will be done in three steps. The first step consists of the derivation of a consistent estimator of ρ and subsequently of φ, on the basis of expression (5.18). In the second step, a consistent estimator of $\mathrm{E}[\xi_i\xi_i']$ is constructed. In the final step, the variance of the corrected estimate of φ needs to be adjusted, to take into account the fact that ρ is unknown and needs to be estimated. Such an adjustment is needed here, since ρ also appears in the bias correction term.

We notice that the nuisance parameter ρ is present in the expression for the bias

of $\hat{\varphi}_{LSDV}$, b_{φ}. Compared to the previous section, the first difficulty is that $\hat{\rho}_{LSDV}$ is inconsistent for ρ. A consistent estimator is obtained in the following way. The probability limit of $\hat{\rho}_{LSDV}$ is $g(\rho) = \rho + b_{\rho}(\rho)$, and is graphed in Figure 5.2. It can be checked empirically that for all T and S, $g(\rho)$ is a monotonically increasing function of ρ, which enables us to compute a corrected estimate for ρ upon inverting g,

$$\hat{\rho}_C = g^{-1}(\hat{\rho}_{LSDV}).$$

This corrected estimate, $\hat{\rho}_C$, is a $\sqrt{N}$-consistent estimator of ρ, since $\hat{\rho}_{LSDV}$ is a $\sqrt{N}$-consistent estimator of $g(\rho) = \rho + b_{\rho}(\rho)$. The first step in obtaining an operational unit root test consists of substituting $\hat{\rho}_C$ for ρ whenever needed, in particular to obtain a bias-corrected estimate of φ,

$$\hat{\varphi}_C = \hat{\varphi}_{LSDV} - b_{\varphi}(\hat{\rho}_C).$$

Whenever $\hat{\rho}_{LSDV}$ exceeds $\lim_{\rho \to 1} g(\rho)$, I propose as an ad hoc solution to set $\hat{\rho}_C = 1$. Similarly, when $\hat{\rho}_{LSDV} < g(-1)$, set $\hat{\rho}_C = -1$. This substitution poses no problems since $b_{\varphi}(1)$ and $b_{\varphi}(-1)$ are defined. More specifically

$$\begin{pmatrix} P(1,T) & Q(1,T) \\ Q(1,T) & R(1,T) \end{pmatrix} = (T-1) \begin{pmatrix} \frac{T\left(T^2+5T-4TS-2+2S\right)}{12} & \frac{T(T+4-3S)}{6} \\ \frac{T(T+4-3S)}{6} & \frac{(T+2-2S)}{2} \end{pmatrix}$$

and

$$\begin{pmatrix} P(-1,T) & Q(-1,T) \\ Q(-1,T) & R(-1,T) \end{pmatrix} = \begin{cases} \begin{pmatrix} \frac{2T^2+T-12-3S(T-2)}{12} & \frac{2T^2+3T-16-4S(T-2)}{8} \\ \frac{2T^2+3T-16-4S(T-2)}{8} & \frac{T^2+T-7-2S(T-2)}{2} \end{pmatrix}, & T \text{ even} \\ \frac{(T-1)(T-3)}{(T-2)} \begin{pmatrix} \frac{(2T-3S+5)}{12} & \frac{(2T-4S+7)}{8} \\ \frac{(2T-4S+7)}{8} & \frac{(T-2S+3)}{2} \end{pmatrix}, & T \text{ odd} \end{cases}$$

are nonsingular matrices for all T and $S \leq 1$.

Another approach could have been to look for another $\sqrt{N}$-consistent estimator for ρ, and to plug this result in $b_{\varphi}(\rho)$. In the context of a stationary dynamic panel, Kiviet (1995) uses a consistent GMM estimate to compute the small-sample distribution of the LSDV-estimator. Harris and Tzavalis (1998) derive a unit root test statistic for panels with finite time dimension, where the errors are heteroskedastic and follow an $MA(1)$ process. The latter test statistic contains the parameter of the $MA(1)$ process as a nuisance parameter, which is removed by estimating the correlation coefficient between Δy_{it} and $\Delta y_{i,t-1}$, and

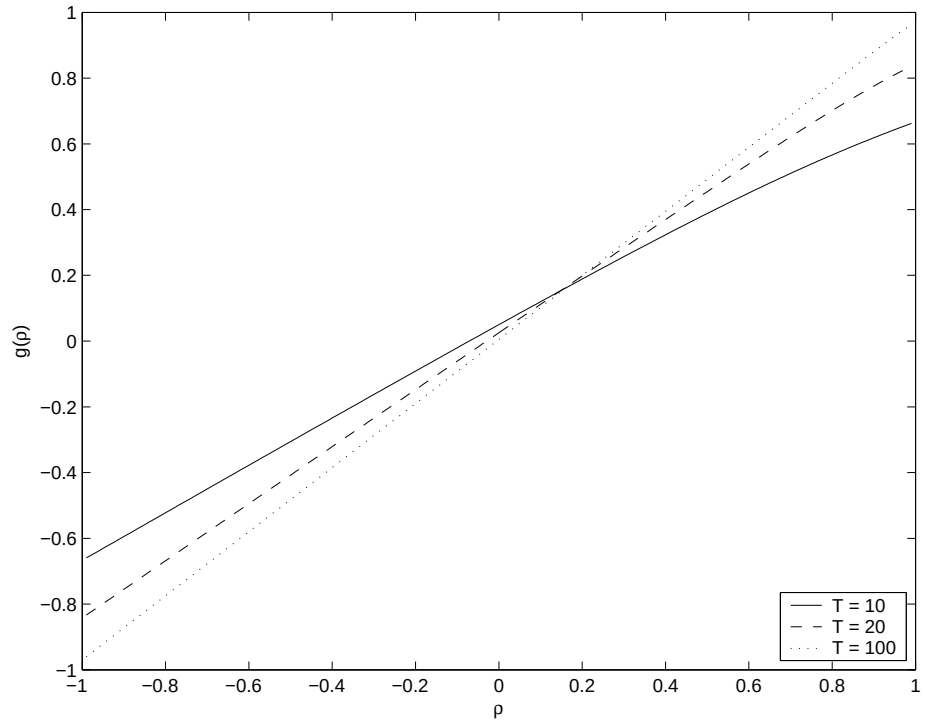

Figure 5.2: The Probability Limit of $\hat{\rho}_{LSDV}$ when $S = 1$.

subsequently inverting the correlation coefficient function of the $MA(1)$ process.

A consistent estimator of $\mathrm{E}[\xi_i \xi_i']$ is obtained as follows. Let

$$\hat{\varepsilon}_i \quad = \quad Q_{T-j} y_i - \hat{\varphi}_C Q_{T-j} y_{i,-1} - \hat{\rho}_C Q_{T-j} \Delta y_{i,-1},$$

then it holds that

$$\hat{\sigma}_\varepsilon^2 \equiv \left[N \mathrm{Tr}\left(Q_{T-j} \right) \right]^{-1} \sum_{i=1}^{N} \hat{\varepsilon}_i' \hat{\varepsilon}_i \quad \xrightarrow{p} \quad \sigma_\varepsilon^2$$

and

$$N^{-1} \sum_{i=1}^{N} \hat{\xi}_i \hat{\xi}_i' \quad \xrightarrow{p} \quad \mathrm{E}\left[\xi_i \xi_i' \right]$$

where $\hat{\xi}_i$ is obtained from (5.19) and (5.13) after replacing φ, ρ, ε_i and σ_ε^2 with $\hat{\varphi}_C$, $\hat{\rho}_C$, $\hat{\varepsilon}_i$ and $\hat{\sigma}_\varepsilon^2$. Estimating these (co)variances has an important advantage compared with using the analytical expressions evaluated at $\hat{\rho}_C$. As with the tests for models with serially

uncorrelated errors (Harris and Tzavalis (1999)), the analytical expression of the variance would involve second and fourth-order moments of the error terms. In order for these moments to cancel out, an assumption about the distribution of the disturbances is needed, or their kurtosis coefficient needs to be estimated. When, on the other hand, the above (co)variances are estimated, no estimation of the kurtosis coefficient of the error terms is needed, since the distribution of the error terms is automatically taken into account.

A final problem is that the variance of the estimator of φ needs to be adjusted, in view of the estimated nature of the bias correction. Indeed,

$$\begin{aligned}\sqrt{N}\left(\hat{\varphi}_C - 1\right) &= \sqrt{N}\left(\hat{\varphi}_{LSDV} - 1 - b_\varphi\left(\hat{\rho}_C\right)\right) \\ &= \sqrt{N}\left(\hat{\varphi}_{LSDV} - 1 - b_\varphi\left(\rho\right)\right) - \sqrt{N}\left(b_\varphi\left(g^{-1}\left(\hat{\rho}_{LSDV}\right)\right) - b_\varphi\left(\rho\right)\right),\end{aligned}$$

which is the sum of two random variables, the second of which can be rewritten as

$$-\sqrt{N}\left[\left(\frac{\partial b_\varphi\left(\rho\right)}{\partial \rho}\right)\left(\frac{\partial g\left(\rho\right)}{\partial \rho}\right)^{-1}\left(\hat{\rho}_{LSDV} - \rho - b_\rho\left(\rho\right)\right)\right] + o_p\left(1\right),$$

since $b_\varphi \circ g^{-1}$ is continuous and differentiable. Since, by Theorem 5.2, $\sqrt{N}\left(\hat{\varphi}, \hat{\rho}\right)'_{LSDV}$ converges in distribution to a bivariate normal distribution, we conclude that $\sqrt{N}\left(\hat{\varphi}_C - 1\right)$ converges in distribution to a normal variate with mean zero and variance given by

$$\sigma^2_{\varphi_c} = G'\Sigma G \qquad \text{with } G = \left(1, -\left(\frac{\partial b_\varphi\left(\rho\right)}{\partial \rho}\right)\left(\frac{\partial g\left(\rho\right)}{\partial \rho}\right)^{-1}\right)'.$$

The unit root test statistic

$$T_2 = \sqrt{N}\frac{\hat{\varphi}_C - 1}{\hat{\sigma}_{\varphi_c}},$$

where $\hat{\sigma}_{\varphi_c}$ is equal to σ_{φ_c} evaluated at $\hat{\varphi}_C$ and $\hat{\rho}_C$, is asymptotically standard normally distributed.

The results of this subsection are specific for the case that $\tau_i = 0$, for all i. Relaxing this assumption would make the test statistic dependent upon the variance of the parameters τ_i. However, by including individual deterministic trends in the regression model, the assumption that $\tau_i = 0$, for all i can be relaxed without introducing extra nuisance parameters, which leads to model 3 in the following subsection.

5.3.3 Model 3

Model 3 consists of the DGP (5.3), and is estimated by the detrended LS dummy variables regression (5.7), whose limiting distribution is presented in the following theorem.

Theorem 5.3. *If Assumptions (A1)-(A3) hold and $\varphi = 1$, then*

$$\sqrt{N}\begin{pmatrix} \hat{\varphi} - 1 - b_{\varphi} \\ \hat{\rho} - \rho - b_{\rho} \end{pmatrix}_{LSDT} \xrightarrow{d} N(0, \Sigma)$$

where

$$\begin{pmatrix} b_{\varphi} \\ b_{\rho} \end{pmatrix} = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1} \begin{pmatrix} V \\ W \end{pmatrix},$$

with

$$\begin{aligned} V &= -A\sum_{t=3}^{T-1}\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho} + B\sum_{t=3}^{T-1}t\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}, \\ W &= -A\sum_{t=3}^{T-1}\sum_{s=3}^{t}\rho^{t-s} + B\sum_{t=3}^{T-1}t\sum_{s=3}^{t}\rho^{t-s} + C\sum_{t=3}^{T-1}\sum_{s=3}^{t}s\rho^{t-s} - D\sum_{t=3}^{T-1}t\sum_{s=3}^{t}s\rho^{t-s}, \end{aligned}$$

$$\begin{aligned} P &= E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^{s}}{1-\rho}\right)^{2} - 2F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{(1-\rho^{r})(1-\rho^{r+t-s})}{(1-\rho)^{2}} \\ &\quad +2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\left(\frac{1-\rho^{s}}{1-\rho}\right)^{2} + 2C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}(t+s)\sum_{r=1}^{s}\frac{(1-\rho^{r})(1-\rho^{r+t-s})}{(1-\rho)^{2}} \\ &\quad -D\sum_{t=2}^{T-1}t^{2}\sum_{s=1}^{t}\left(\frac{1-\rho^{s}}{1-\rho}\right)^{2} - 2D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{(1-\rho^{r})(1-\rho^{r+t-s})}{(1-\rho)^{2}} \\ &\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^{t}}{1-\rho}\right)^{2} - \omega F\left(\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\right)^{2} \\ &\quad +2\omega C\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}\frac{1-\rho^{s}}{1-\rho} - \omega D\left(\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\right)^{2}, \end{aligned}$$

$$
\begin{aligned}
Q &= E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}-F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&+2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}+C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}(t+s)\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&-D\sum_{t=2}^{T-1}t^{2}\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}-D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&+\omega\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\rho^{t-1}-\omega F\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}+\omega C\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}\\
&+\omega C\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1}-\omega D\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1},\\
R &= E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2}-2F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2}\\
&+2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\rho^{2s-2}+2C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}(t+s)\sum_{r=1}^{s}\rho^{2r+t-s-2}\\
&-D\sum_{t=2}^{T-1}t^{2}\sum_{s=1}^{t}\rho^{2s-2}-2D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{2r+t-s-2}\\
&+\omega\sum_{t=2}^{T-1}\rho^{2t-2}-\omega F\left(\sum_{t=2}^{T-1}\rho^{t-1}\right)^{2}\\
&+2\omega C\sum_{t=2}^{T-1}t\rho^{t-1}\sum_{s=2}^{T-1}\rho^{s-1}-\omega D\left(\sum_{t=2}^{T-1}t\rho^{t-1}\right)^{2}
\end{aligned}
$$

and

$$
\begin{aligned}
A &= \tfrac{4\left(T^{2}+2T+3\right)}{(T-1)(T-2)(T-3)}, & B &= \tfrac{6(T+3)}{(T-1)(T-2)(T-3)},\\
C &= \tfrac{6(T+1)}{(T-1)(T-2)(T-3)}, & D &= \tfrac{12}{(T-1)(T-2)(T-3)},\\
E &= \tfrac{T^{3}-10T^{2}+9T-12}{(T-1)(T-2)(T-3)}, & F &= \tfrac{2\left(2T^{2}+T+3\right)}{(T-1)(T-2)(T-3)},
\end{aligned}
$$

and where

$$
\Sigma = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1} E[\xi_i\xi_i'] \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}^{-1},
$$

with

$$\xi_i = \begin{pmatrix} v_i \\ w_i \end{pmatrix} - \begin{pmatrix} p_i & q_i \\ q_i & r_i \end{pmatrix} \begin{pmatrix} b_\varphi \\ b_\rho \end{pmatrix},$$

were p_i, q_i, r_i, v_i and w_i are defined in (5.13).

The procedure to obtain an operational test statistic is identical to the method outlined under Model 2. Defining the function $h(\rho)$ as the probability limit of $\hat{\rho}_{LSDT}$, i.e. $h(\rho) = \rho + b_\rho(\rho)$, a corrected estimate for ρ is obtained by inverting h,

$$\hat{\rho}_C = h^{-1}(\hat{\rho}_{LSDT})$$

and a bias-corrected estimate of φ is given by

$$\hat{\varphi}_C = \hat{\varphi}_{LSDT} - b_\varphi(\hat{\rho}_C).$$

Following the same reasoning as for Model 2, $\sqrt{N}(\hat{\varphi}_C - 1)$ converges in distribution to a normal variate with mean zero and variance given by

$$\sigma^2_{\varphi_c} = H'\Sigma H \qquad \text{with } H = \left(1, -\left(\frac{\partial b_\varphi(\rho)}{\partial \rho}\right)\left(\frac{\partial h(\rho)}{\partial \rho}\right)^{-1}\right)'.$$

Consequently, the unit root test statistic

$$T_3 = \sqrt{N}\frac{\hat{\varphi}_C - 1}{\hat{\sigma}_{\varphi_c}},$$

where $\hat{\sigma}_{\varphi_c}$ is equal to σ_{φ_c} evaluated at $\hat{\varphi}_C$ and $\hat{\rho}_C$, is asymptotically standard normally distributed.

One point needs to be clarified, however. In Figure 5.3, $h(\rho)$ is plotted. For $T > 5$, it is no longer a monotonic function of ρ, but reaches a maximum at some point ρ_M inside the interval $(-1, 1)$. The location of this maximum increases for increasing values of T and S. These maximum values, h_M, and the points where they are reached, ρ_M, are tabulated for selected values of T for the error process starting in period 1 in Table 5.1 and for the error process starting in the infinite past in Table 5.2.

The nonmonotonicity of $h(\rho)$ raises two problems for our test statistic. First, for some values of $\hat{\rho}_{LSDT}$, we need to choose between two possible values of $\hat{\rho}_C$. Second, in ρ_M

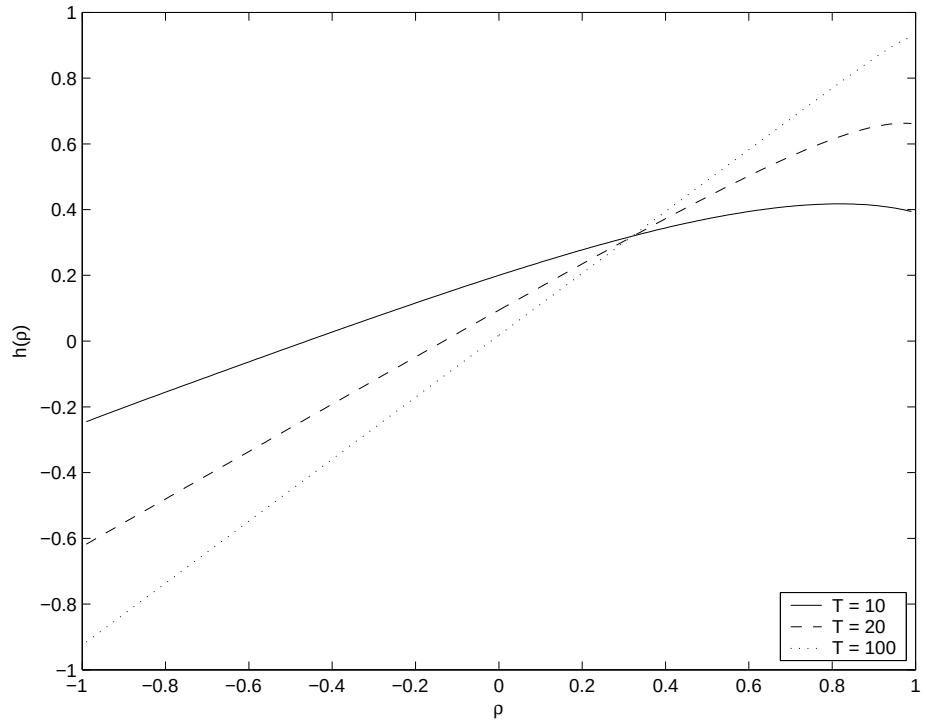

Figure 5.3: The Probability Limit of $\hat{\rho}_{LSDT}$ when $S = 1$.

our test statistic has infinite variance. The following ad hoc solution which tackles both problems, stems from the observation that, for $T = 5$, the funcion $h(\rho)$ is monotonous. Define h_D as the minimum value of $h(\rho)$, for $\rho \in [-1, 1]$, for which one of the two problems occurs. It is equal to either $h(-1)$ or $h(1)$. Whenever, for the number of periods T at hand, $\hat{\rho}_{LSDT;T} \geq h_{D;T}$, re-estimate ρ using only five consecutive periods, which results in $(\hat{\varphi}, \hat{\rho})_{LSDT;5}$. A corrected estimate of ρ is obtained by inverting $h(\rho; 5)$,

$$\hat{\rho}_D = h^{-1}(\hat{\rho}_{LSDT}; 5).$$

The bias-corrected estimate of φ will be computed as

$$\hat{\varphi}_D = \hat{\varphi}_{LSDT;T} - b_\varphi(\hat{\rho}_D; T).$$

$\sqrt{N}(\hat{\varphi}_D - 1)$ converges in distribution to a normal variate with mean zero and variance

T	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
ρ_M	-0.20866	0.51848	0.74218	0.80298	0.83857	0.86309	0.88131	0.89547
$\mathbf{h}_M$	0.33775	0.32111	0.36081	0.40223	0.44063	0.47548	0.50690	0.53520
T	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**25**
ρ_M	0.90682	0.91614	0.92392	0.93052	0.93617	0.94107	0.94535	0.96053
$\mathbf{h}_M$	0.56072	0.58378	0.60469	0.62370	0.64104	0.65692	0.67150	0.72939
T	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
ρ_M	0.96967	0.97995	0.98543	0.98877	0.99099	0.99255	0.99370	0.99457
$\mathbf{h}_M$	0.77019	0.82365	0.85704	0.87983	0.89638	0.90892	0.91876	0.92669

Table 5.1: The maxima of $\hat{\rho}_{LSDT}(\rho, T)$ for different values of T when $S = 1$.

given by

$$\sigma^2_{\varphi_D} = \tilde{H}'\tilde{\Sigma}\tilde{H} \qquad \text{with } \tilde{H} = \left(1, -\left(\frac{\partial b_\varphi(\rho;T)}{\partial \rho}\right)\left(\frac{\partial h(\rho;5)}{\partial \rho}\right)^{-1}\right)',$$

where $\tilde{\Sigma}$ is the variance matrix of $(\hat{\varphi}_T, \hat{\rho}_5)_{LSDT}$, which is as straightforward to estimate as Σ. As in model 2, when $\hat{\rho}_{LSDV} < g(-1)$, set $\hat{\rho}_C = -1$. This substitution is possible since $b_\varphi(-1)$ is defined because

$$\begin{pmatrix} P(-1,T) & Q(-1,T) \\ Q(-1,T) & R(-1,T) \end{pmatrix}$$

$$= \begin{cases} \frac{T(T-4)}{(T-1)(T-3)} \begin{pmatrix} \frac{\left(17T^2+22T-114-30S(T-2)\right)}{120} & \frac{(T-2)(2T+7-4S)}{8} \\ \frac{(T-2)(2T+7-4S)}{8} & \frac{(T-2)(T+3-2S)}{2} \end{pmatrix}, & T \text{ even} \\ \begin{pmatrix} \frac{17T^3-12T^2-209T+240-30S(T-1)(T-3)}{120(T-2)} & \frac{2T^3-T^2-26T+29-4S(T-1)(T-3)}{8(T-2)} \\ \frac{2T^3-T^2-26T+29-4S(T-1)(T-3)}{8(T-2)} & \frac{T^3-T^2-11T+13-2S(T-1)(T-3)}{2(T-2)} \end{pmatrix}, & T \text{ odd} \end{cases}$$

is nonsingular.

5.4 Concluding Remarks

This chapter introduces asymptotic unit root tests for panel data models where the errors are serially correlated and the time dimension is fixed. The tests allow for fixed effects and individual deterministic time trends, and are based on the standardized least squares estimator of the autoregressive coefficient. When fixed effects and/or individual time trends are included in the regression model, the coefficient is estimated inconsistently by standard

T	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
ρ_M	-0.19137	0.37871	0.62083	0.74554	0.81771	0.86310	0.89346	0.91475
$\mathbf{h}_M$	0.33757	0.31729	0.34273	0.37942	0.41748	0.45366	0.48694	0.51718
T	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**25**
ρ_M	0.93025	0.94188	0.95083	0.95786	0.96349	0.96806	0.97182	0.98343
$\mathbf{h}_M$	0.54453	0.56925	0.59164	0.61195	0.63044	0.64732	0.66279	0.72378
T	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
ρ_M	0.98910	0.99426	0.99647	0.99761	0.99827	0.99870	0.99898	0.99918
$\mathbf{h}_M$	0.76634	0.82157	0.85576	0.87898	0.89577	0.90848	0.91842	0.92642

Table 5.2: The maxima of $\hat{\rho}_{LSDT}(\rho, T)$ for different values of T when $S = -\infty$.

estimators and a bias-correction is needed. The limiting distributions of the tests are shown to be normal. The inconsistency is calculated analytically and a procedure is proposed for the estimation of the variance. A comparison is made with the tests proposed by Harris and Tzavalis (1999), and it is shown that our test performs better, even if the amount of serial correlation is fairly small.

Appendix A

A Simple Estimator: Proofs of Theorems

A.1 Decomposition of $h_{\varepsilon;i}$

Recalling that the vector r_i is defined as $\text{Vech}\left((1, x_i', w_i', q_i')'(1, x_i', w_i', q_i')\right)'$, the function $h_{\varepsilon;i}$ from (2.7) can be written as a polynomial of second degree in (x_i, w_i, q_i, v_i) plus a remainder, $\mathcal{R}_{\varepsilon;i}$:

$$h_{\varepsilon;i} = r_i'\lambda_1 + \left(v_i', \text{Vec}\left(x_i v_i'\right)', \text{Vec}\left(w_i v_i'\right)', \text{Vec}\left(q_i v_i'\right)', \text{Vech}\left(v_i v_i'\right)'\right)\lambda_2 + \mathcal{R}_{\varepsilon;i},$$

where $\text{E}\left[\mathcal{R}_{\varepsilon;i}\mathcal{P}_2\left(x_i, w_i, q_i, v_i\right)\right] = 0$, with $\mathcal{P}_2\left(x_i, w_i, q_i, v_i\right)$ any polynomial of maximal degree two. Making use of the definitions of e_i and v_i^*, repeated here for convenience,

$$\begin{aligned} e_i &= \text{Vech}\left(v_i v_i'\right) - \text{E}\left[\text{Vech}\left(v_i v_i'\right) r_i'\right]\left(\text{E}\left[r_i r_i'\right]\right)^{-1} r_i^* \\ v_i^* &= \left(v_i', \text{Vec}\left(x_i v_i'\right)', \text{Vec}\left(w_i v_i'\right)', \text{Vec}\left(q_i v_i'\right)', e_i'\right)', \end{aligned}$$

$h_{\varepsilon;i}$ can be re-parametrized as

$$h_{\varepsilon;i} = \kappa' r_i + \theta_\varepsilon' v_i^* + \mathcal{R}_{\varepsilon;i}. \tag{A.1}$$

From Assumption A and from (2.7) follows that

$$\text{E}\left[\varepsilon_i \mid x_i, w_i, q_i\right] = \text{E}\left[\text{E}\left[\varepsilon_i \mid x_i, w_i, q_i, v_i\right] \mid x_i, w_i, q_i\right] = \text{E}\left[h_{\varepsilon;i} \mid x_i, w_i, q_i\right] = 0 \ .$$

Premultiplying both sides of (A.1) with r_i, we deduce that $\kappa = 0$. On the left-hand side $\mathrm{E}[r_i h_{\varepsilon;i}] = 0$, since $\mathrm{E}[h_{\varepsilon;i} \mid x_i, w_i, q_i] = 0$. On the right-hand side it holds that $\mathrm{E}[r_i \mathcal{R}_{\varepsilon;i}] = 0$, since the elements of r_i are monomials in x_i, w_i, q_i and v_i of degree two or less, and we also have that $\mathrm{E}[r_i v_i'] = 0$, by Assumption A and by definition of e_i.

A.2 Inconsistency of the SB Estimator

The inconsistency of the SB estimator is proved for scalar z_i, which simplifies (2.2), (2.10) and (2.11) to

$$\begin{aligned} z_i &= \beta_0 + \beta_1 x_i + \beta_2 w_i + v_i, \\ e_i &= v_i^2 - r_i'\delta \\ h_{\varepsilon;i} &= v_i\theta_{\varepsilon 1} + v_i x_i'\theta_{\varepsilon 2} + v_i w_i'\theta_{\varepsilon 3} + v_i q_i'\theta_{\varepsilon 4} + e_i\theta_{\varepsilon 5} + \mathcal{R}_{\varepsilon;i}. \end{aligned}$$

I will prove that

$$f_{\varepsilon;i} = h_{\varepsilon;i} - \mathrm{E}\left[(v_i h_{\varepsilon;i}, z_i v_i h_{\varepsilon;i})\right] \left(\mathrm{E} \begin{bmatrix} v_i^2 & z_i v_i^2 \\ z_i v_i^2 & z_i^2 v_i^2 \end{bmatrix} \right)^{-1} \begin{pmatrix} v_i \\ z_i v_i \end{pmatrix} \tag{A.2}$$

is correlated with (x_i', z_i), which implies inconsistency of Garen's (1984) SB estimator of the outcome equation (2.6). This estimator is given by

$$\hat{\alpha}_{SB} = \left(\sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ z_i^* \\ z_i v_i \\ v_i \end{pmatrix} (1, x_i, z_i^*, z_i v_i, v_i) \right)^{-1} \sum_{i=1}^{N} \begin{pmatrix} 1 \\ x_i \\ z_i^* \\ z_i v_i \\ v_i \end{pmatrix} y_i.$$

In order to prove it's incosistency, I compute the projection of $h_{\varepsilon;i}$ on $(v_i, z_i v_i)'$, i.e. the second term of (A.2). Taking into account the following equalities

$$\begin{aligned} \mathrm{E}[v_i h_{\varepsilon;i}] &= \mathrm{E}\left[v_i^2\right]\theta_{\varepsilon 1} + \mathrm{E}\left[v_i^2 x_i'\right]\theta_{\varepsilon 2} + \mathrm{E}\left[v_i^2 w_i'\right]\theta_{\varepsilon 3} \\ &\quad + \mathrm{E}\left[v_i^2 q_i'\right]\theta_{\varepsilon 4} + \mathrm{E}\left[v_i e_i\right]\theta_{\varepsilon 5} \end{aligned}$$

$$\begin{aligned}\mathrm{E}\left[z_i v_i h_{\varepsilon;i}\right] &= \mathrm{E}\left[z_i v_i^2\right]\theta_{\varepsilon 1} + \mathrm{E}\left[z_i v_i^2 x_i'\right]\theta_{\varepsilon 2} + \mathrm{E}\left[z_i v_i^2 w_i'\right]\theta_{\varepsilon 3} \\ &\quad + \mathrm{E}\left[z_i v_i^2 q_i'\right]\theta_{\varepsilon 4} + \mathrm{E}\left[z_i v_i e_i\right]\theta_{\varepsilon 5}\end{aligned}$$

$$\begin{aligned}&\left(\mathrm{E}\begin{bmatrix} v_i^2 & z_i v_i^2 \\ z_i v_i^2 & z_i^2 v_i^2 \end{bmatrix}\right)^{-1} \\ &= \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i^2 v_i^2\right] - \left(\mathrm{E}\left[z_i v_i^2\right]\right)^2\right)^{-1} \begin{pmatrix} \mathrm{E}\left[z_i^2 v_i^2\right] & -\mathrm{E}\left[z_i v_i^2\right] \\ -\mathrm{E}\left[z_i v_i^2\right] & \mathrm{E}\left[v_i^2\right] \end{pmatrix},\end{aligned}$$

we have that

$$\begin{aligned}&\mathrm{E}\left[(v_i h_{\varepsilon;i}, z_i v_i h_{\varepsilon;i})\right]\left(\mathrm{E}\begin{bmatrix} v_i^2 & z_i v_i^2 \\ z_i v_i^2 & z_i^2 v_i^2 \end{bmatrix}\right)^{-1}\begin{pmatrix} v_i \\ z_i v_i \end{pmatrix} \\ &= \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i^2 v_i^2\right] - \left(\mathrm{E}\left[z_i v_i^2\right]\right)^2\right)^{-1} \\ &\quad \times \begin{pmatrix} \mathrm{E}\left[v_i h_{\varepsilon;i}\right] & \mathrm{E}\left[z_i v_i h_{\varepsilon;i}\right] \end{pmatrix}\begin{pmatrix} \mathrm{E}\left[z_i^2 v_i^2\right] & -\mathrm{E}\left[z_i v_i^2\right] \\ -\mathrm{E}\left[z_i v_i^2\right] & \mathrm{E}\left[v_i^2\right] \end{pmatrix}\begin{pmatrix} v_i \\ z_i v_i \end{pmatrix} \\ &= \frac{v_i A + z_i v_i B}{N},\end{aligned}$$

where

$$\begin{aligned}A &= \mathrm{E}\left[v_i h_{\varepsilon;i}\right]\mathrm{E}\left[z_i^2 v_i^2\right] - \mathrm{E}\left[z_i v_i h_{\varepsilon;i}\right]\mathrm{E}\left[z_i v_i^2\right] \\ &= N\theta_{\varepsilon 1} \\ &\quad + \left(\mathrm{E}\left[z_i^2 v_i^2\right]\mathrm{E}\left[v_i^2 x_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[z_i v_i^2 x_i'\right]\right)\theta_{\varepsilon 2} \\ &\quad + \left(\mathrm{E}\left[z_i^2 v_i^2\right]\mathrm{E}\left[v_i^2 w_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[z_i v_i^2 w_i'\right]\right)\theta_{\varepsilon 2} \\ &\quad + \left(\mathrm{E}\left[z_i^2 v_i^2\right]\mathrm{E}\left[v_i^2 q_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[z_i v_i^2 q_i'\right]\right)\theta_{\varepsilon 2} \\ &\quad + \left(\mathrm{E}\left[z_i^2 v_i^2\right]\mathrm{E}\left[v_i e_i\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[z_i v_i e_i\right]\right)\theta_{\varepsilon 3},\end{aligned}$$

$$\begin{aligned}B &= -\mathrm{E}\left[v_i h_{\varepsilon;i}\right]\mathrm{E}\left[z_i v_i^2\right] + \mathrm{E}\left[z_i v_i h_{\varepsilon;i}\right]\mathrm{E}\left[v_i^2\right] \\ &= \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i v_i^2 x_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[v_i^2 x_i'\right]\right)\theta_{\varepsilon 2} \\ &\quad + \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i v_i^2 w_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[v_i^2 w_i'\right]\right)\theta_{\varepsilon 3} \\ &\quad + \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i v_i^2 q_i'\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[v_i^2 q_i'\right]\right)\theta_{\varepsilon 4} \\ &\quad + \left(\mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i v_i e_i\right] - \mathrm{E}\left[z_i v_i^2\right]\mathrm{E}\left[v_i e_i\right]\right)\theta_{\varepsilon 5}\end{aligned}$$

and

$$N = \mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i^2 v_i^2\right] - \left(\mathrm{E}\left[z_i v_i^2\right]\right)^2.$$

Expression (A.2) consequently becomes equal to

$$\begin{aligned} f_{\varepsilon;i} &= v_i\left(\theta_{\varepsilon 1} - \frac{A}{N}\right) + v_i x_i'\theta_{\varepsilon 2} + v_i w_i'\theta_{\varepsilon 3} \\ &\quad + v_i q_i'\theta_{\varepsilon 4} + e_i\theta_{\varepsilon 5} - v_i z_i\frac{B}{N} + \mathcal{R}_{\varepsilon;i}, \end{aligned}$$

and it holds that

$$\mathrm{E}\left[x_i f_{\varepsilon;i}\right] = -N^{-1}\mathrm{E}\left[x_i v_i^2\right] B \tag{A.3}$$

$$\begin{aligned} \mathrm{E}\left[z_i f_{\varepsilon;i}\right] &= N^{-1}\left\{\mathrm{E}\left[z_i v_i\right](N\theta_{\varepsilon 1} - A) + \mathrm{E}\left[z_i v_i x_i'\right] N\theta_{\varepsilon 2} + \mathrm{E}\left[z_i v_i w_i'\right] N\theta_{\varepsilon 3}\right. \\ &\quad \left. + \mathrm{E}\left[z_i v_i q_i'\right] N\theta_{\varepsilon 4} + \mathrm{E}\left[z_i e_i\right] N\theta_{\varepsilon 5} - \mathrm{E}\left[v_i z_i^2\right] B\right\} \\ &= N^{-1}\left\{\mathrm{E}\left[v_i^2\right](N\theta_{\varepsilon 1} - A) + \mathrm{E}\left[v_i^2 x_i'\right] N\theta_{\varepsilon 2} + \mathrm{E}\left[v_i^2 w_i'\right] N\theta_{\varepsilon 3}\right. \\ &\quad \left. + \mathrm{E}\left[v_i^2 q_i'\right] N\theta_{\varepsilon 4} + \mathrm{E}\left[v_i e_i\right] N\theta_{\varepsilon 5} - \mathrm{E}\left[v_i z_i^2\right] B\right\} \\ &= \mathrm{E}\left[v_i^2\left(1, x_i', w_i'\right)\beta\right]\left(\mathrm{E}\left[v_i^2 x_i'\right]\mathrm{E}\left[z_i v_i^2\right] - \mathrm{E}\left[z_i v_i^2 x_i'\right]\mathrm{E}\left[v_i^2\right]\right)\theta_{\varepsilon 2} \\ &\quad + \mathrm{E}\left[v_i^2\left(1, x_i', w_i'\right)\beta\right]\left(\mathrm{E}\left[v_i^2 w_i'\right]\mathrm{E}\left[z_i v_i^2\right] - \mathrm{E}\left[z_i v_i^2 w_i'\right]\mathrm{E}\left[v_i^2\right]\right)\theta_{\varepsilon 3} \\ &\quad + \mathrm{E}\left[v_i^2\left(1, x_i', w_i'\right)\beta\right]\left(\mathrm{E}\left[v_i^2 q_i'\right]\mathrm{E}\left[z_i v_i^2\right] - \mathrm{E}\left[z_i v_i^2 q_i'\right]\mathrm{E}\left[v_i^2\right]\right)\theta_{\varepsilon 4} \\ &\quad + \mathrm{E}\left[v_i^2\left(1, x_i', w_i'\right)\beta\right]\left(\mathrm{E}\left[v_i e_i\right]\mathrm{E}\left[z_i v_i^2\right] - \mathrm{E}\left[v_i^2\right]\mathrm{E}\left[z_i v_i e_i\right]\right)\theta_{\varepsilon 5}. \end{aligned} \tag{A.4}$$

In general, neither (A.3) nor (A.4) is equal to zero. In the following cases, however, Garen's (1984) SB estimator is consistent:

- The linearity Assumption (2.4) holds, which implies that $\theta_{\varepsilon 2}, \ldots, \theta_{\varepsilon 5}$ is equal to zero.
- The second and third order conditional moments of the disturbances of equation (2.2) are not a function of the exogenous variables (x_i', w_i', q_i') in the model:

$$\begin{aligned} \mathrm{E}\left[v_i^2 \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[v_i^2\right], \\ \mathrm{E}\left[v_i^3 \mid x_i, w_i, q_i\right] &= \mathrm{E}\left[v_i^3\right]. \end{aligned}$$

In both cases it holds that (A.3) and (A.4) are both equal to zero.

A.3 Asymptotic Normality

Lemma A.1. *It holds that*

$$N^{-\frac{1}{2}}\sum_{i=1}^{N}\hat{s}_i\tilde{\zeta}_i \quad \xrightarrow{d} \quad N\left(0, E\left[\zeta_i^2 s_i s_i'\right] + D\Omega D'\right).$$

Proof. In a first step, expression (2.17) for $\mathrm{E}\left[s_i\left(\frac{\partial\tilde{\zeta}_i}{\partial\beta'}\right)\right]$ is proved, next asymptotic normality is proved.

The disturbance (2.15), can be rewritten as

$$\begin{aligned}\tilde{\zeta}_i &= \zeta_i - \theta_1'\left\{I_L\otimes(1,x_i',w_i')\right\}\left(\beta-\hat{\beta}\right) - \theta_2'\mathrm{Vec}\left(x_i\left(\beta-\hat{\beta}\right)'\left\{I_L\otimes(1,x_i',w_i')'\right\}\right)\\ &\quad -\theta_3'\mathrm{Vec}\left(w_i\left(\beta-\hat{\beta}\right)'\left\{I_L\otimes(1,x_i',w_i')'\right\}\right) - \theta_4'\mathrm{Vec}\left(q_i\left(\beta-\hat{\beta}\right)'\left\{I_L\otimes(1,x_i',w_i')'\right\}\right)\\ &\quad +\theta_5'\mathrm{Vech}\left(v_iv_i' - \hat{v}_i\hat{v}_i'\right) - \theta_6 z_i'\left\{I_L\otimes(1,x_i',w_i')\right\}\left(\beta-\hat{\beta}\right),\end{aligned}$$

since

$$\begin{aligned}v_i-\hat{v}_i &= -\left(\beta_0-\hat{\beta}_0\right) - \left(B_1-\hat{B}_1\right)x_i - \left(B_2-\hat{B}_2\right)w_i\\ &= -\left\{I_L\otimes(1,x_i',w_i')\right\}\left(\beta-\hat{\beta}\right)\\ &= -\left(\beta-\hat{\beta}\right)'\left\{I_L\otimes(1,x_i',w_i')'\right\}.\end{aligned}$$

It is of the form $\zeta_i + f(\beta) - f\left(\hat{\beta}\right)$, and can asymptotically be approximated by

$$\tilde{\zeta}_i \;=\; \zeta_i - \left(\frac{\partial f}{\partial\beta'}\right)\left(\hat{\beta}-\beta\right) + o_p(1).$$

Since

$$\left(\partial\mathrm{Vec}\left((1,x_i',w_i',q_i')'\beta'\left\{I_L\otimes(1,x_i',w_i')'\right\}\right)/\partial\beta'\right) \;=\; \left\{I_L\otimes(1,x_i',w_i',q_i')'(1,x_i',w_i')\right\}$$

and

$$\left(\partial\mathrm{Vec}\left(v_iv_i'\right)/\partial\beta'\right) \;=\; \left\{\Phi_i\otimes(1,x_i',w_i')\right\},$$

with $\Phi_i = (\varphi_{i;1}, \ldots, \varphi_{i;L})$, $\varphi_{i;j} = \text{Vech}\left(\mathbf{1}_{j;L} v_i' + v_i \mathbf{1}_{j;L}'\right)$ and $\mathbf{1}_{j;L}$ the j-th $1 \times L$ unit vector, we have that

$$\begin{aligned}\frac{\partial f}{\partial \beta'} &= -\theta_1' \left\{I_L \otimes (1, x_i', w_i')\right\} - \theta_2' \left\{I_L \otimes x_i \left(1, x_i', w_i'\right)\right\} \\ &\quad -\theta_3' \left\{I_L \otimes w_i \left(1, x_i', w_i'\right)\right\} - \theta_4' \left\{I_L \otimes q_i \left(1, x_i', w_i'\right)\right\} \\ &\quad -\theta_5' \left\{\Phi_i \otimes (1, x_i', w_i')\right\} - \theta_6' \left\{z_i' \otimes (1, x_i', w_i')\right\}. \end{aligned} \tag{A.5}$$

Premultiplying (A.5) with s_i and taking expectations results in (2.17)

$$\begin{aligned} D &= \text{E}\left[s_i \left(\frac{\partial \tilde{\zeta}_i}{\partial \beta'}\right)\right] = \text{E}\left[-s_i \left(\frac{\partial f}{\partial \beta'}\right)\right] \\ &= \text{E}\left[s_i \left(\theta_1' \left\{I_L \otimes (1, x_i', w_i')\right\} + \theta_2' \left\{I_L \otimes \left(x_i \left(1, x_i', w_i'\right)\right)\right\}\right)\right] \\ &\quad +\text{E}\left[s_i \left(\theta_3' \left\{I_L \otimes \left(w_i \left(1, x_i', w_i'\right)\right)\right\} + \theta_4' \left\{I_L \otimes \left(q_i \left(1, x_i', w_i'\right)\right)\right\}\right)\right] \\ &\quad +\text{E}\left[s_i \left(\theta_5' \left\{\Phi_i \otimes (1, x_i', w_i')\right\} + \theta_6' \left(z_i' \otimes \left(1, x_i', w_i'\right)\right)\right)\right]. \end{aligned}$$

We have, by Assumption A-C, and by the definition of D in (2.17), that

$$\begin{aligned} N^{-\frac{1}{2}} \sum_{i=1}^{N} \hat{s}_i \tilde{\zeta}_i &= N^{-\frac{1}{2}} \sum_{i=1}^{N} \hat{s}_i \zeta_i + N^{-\frac{1}{2}} \sum_{i=1}^{N} \Big\{\hat{s}_i \left(v_i - \hat{v}_i\right)' \theta_1 \\ &\quad +\text{Vec}\left(x_i \left(v_i - \hat{v}_i\right)'\right)' \theta_2 + \text{Vec}\left(w_i \left(v_i - \hat{v}_i\right)'\right)' \theta_3 \\ &\quad +\text{Vec}\left(q_i \left(v_i - \hat{v}_i\right)'\right)' \theta_4 + \left(e_i - \hat{e}_i\right)' \theta_5 + z_i' \left(v_i - \hat{v}_i\right) \theta_6 \Big\} \\ &= N^{-\frac{1}{2}} \sum_{i=1}^{N} s_i \zeta_i + D\left[\sqrt{N}\left(\hat{\beta} - \beta\right)\right] + o_p(1) \\ &= N^{-\frac{1}{2}} \sum_{i=1}^{N} s_i \zeta_i + D\text{Vec}\left(N^{-\frac{1}{2}} \sum_{i=1}^{N} \delta_i\right) + o_p(1). \end{aligned}$$

The first two terms on the right-hand side are independent processes, which are uncorrelated, with variances equal to $\text{E}\left[s_i \zeta_i^2 s_i'\right]$, respectively $D\Omega D'$. The conclusion follows by the Lyapunov CLT. □

Appendix B

Semiparametric Estimation: Proofs of Theorems

B.1 Preliminaries

Definitions. Define the $N \times J$, respectively $N \times L$ matrices

$$\begin{aligned} H_x &= (h_x(\tilde{v}_1), \ldots, h_x(\tilde{v}_N))', \\ H_z^* &= (z_1 h_z(\tilde{v}_1), \ldots, z_N h_z(\tilde{v}_N))', \end{aligned}$$

and also $H^* = (H_x, H_z^*)$. Define in the same fashion $G^* = G_\varepsilon + G_u^*$.

The first $\kappa_m(N) = \max(\kappa_\varepsilon(N), \kappa_u(N))$ elements of the family of orthonormal polynomials with respect to the distribution $F(\tilde{v})$, $\{\varphi_j(\cdot) : j = 1, 2, \ldots\}$, will be estimated by

$$\hat{\vec{\varphi}}_{\kappa_m}(\cdot) = (\hat{\varphi}_{1N}(\cdot), \ldots, \hat{\varphi}_{\kappa_m N}(\cdot))',$$

where

$$\int_{-\infty}^{\infty} \hat{\varphi}_{jN}\left(\hat{\tilde{v}}\right) \hat{\varphi}_{kN}\left(\hat{\tilde{v}}\right) d\hat{F}_N\left(\hat{\tilde{v}}\right) = N^{-1} \sum_{i=1}^{N} \hat{\varphi}_{jN}\left(\hat{\tilde{v}}_i\right) \hat{\varphi}_{kN}\left(\hat{\tilde{v}}_i\right) = \delta_{jk},$$

with $\hat{F}_N$ the empirical CDF of $\hat{\tilde{v}}$ and δ_{jk} Kronecker's delta. The following lemma assures the convergence of the estimated orthogonal polynomials to the true orthogonal polynomials.

Lemma B.1. $\hat{\varphi}_{jN}\left(\cdot; \hat{\beta}\right) \xrightarrow{p} \varphi_j(\cdot; \beta), \quad \forall j = 1, \ldots, \kappa_m(N).$

Proof. Since $\tilde{v}_i = \mathrm{E}_\beta [v_i | x_i, w_i, z_i]$, we can rewrite $\varphi_j(\cdot)$ as $\varphi_j(\cdot;\beta)$ which will be approximated by $\hat{\varphi}_{jN}(\cdot) = \hat{\varphi}_{jN}\left(\cdot;\hat{\beta}\right)$. Now, since (i) $\hat{\beta} \xrightarrow{p} \beta$ and since (ii) $\hat{\varphi}_{jN}(\cdot;\theta) \xrightarrow{p} \varphi_j(\cdot;\theta)$ uniformly on an open set B_0 containing $\theta = \beta$, with $\varphi_j(z,p;\theta)$ continuous at $\theta = \beta$, it holds that $\hat{\varphi}_{j\kappa}\left(.;\hat{\beta}\right) \xrightarrow{p} \varphi_j(.;\beta)$ (Davidson (1994), theorem 21.6).

Condition (i) is satisfied by Assumption C and condition (ii) follows immediately from Theorem 2 by Jennrich (1969). □

Lemma B.2. *For any $N \times I$-matrix A and $N \times J$-matrix B for which $N^{-1}A'A = O_p(1)$ and $N^{-1}B'B = O_p(1)$, it holds that*

$$N^{-1}A'\left(\mathcal{P}_\Pi - \mathcal{P}_{\hat{\Pi}}\right)B \;=\; O_p\left(N^{-\frac{1}{2}}\right),$$

where $\hat{\Pi}$ is a $N \times \kappa$ matrix of powers in $\tilde{v}_i, \ldots, \tilde{v}_i^{(\kappa+1)}$. The variables $\tilde{v}_i$ are continuous functions of some estimator $\hat{\beta}$ for which it holds that $\sqrt{N}\left(\hat{\beta} - \beta\right) = O_p(1)$. If, in addition, it holds that $N^{-1}A'\Pi = o_p(1)$, or $N^{-1}B'\Pi = o_p(1)$ then it holds that

$$N^{-1}A'\left(\mathcal{P}_\Pi - \mathcal{P}_{\hat{\Pi}}\right)B \;=\; o_p\left(N^{-\frac{1}{2}}\right).$$

Proof. Rewrite the left-hand side as

$$\begin{aligned}
N^{-1}A'\left(\mathcal{P}_\Pi - \mathcal{P}_{\hat{\Pi}}\right)B \;=\;& N^{-1}A'\Pi\left(\Pi'\Pi\right)^{-1}\Pi'B \\
& -N^{-1}A'\hat{\Pi}\left(\hat{\Pi}'\hat{\Pi}\right)^{-1}\hat{\Pi}'B \\
=\;& \left[N^{-1}A'\left(\hat{\Pi}-\Pi\right)\right]\left(N^{-1}\Pi'\Pi\right)^{-1}\left[N^{-1}\hat{\Pi}'B\right] \\
& + \left[N^{-1}A'\Pi\right]\left(N^{-1}\Pi'\Pi\right)^{-1}\left[N^{-1}\left(\Pi-\hat{\Pi}\right)'B\right] \\
& + \left[N^{-1}A'\hat{\Pi}\right]N\left\{\left(\Pi'\Pi\right)^{-1} - \left(\hat{\Pi}'\hat{\Pi}\right)^{-1}\right\}\left[N^{-1}\hat{\Pi}'B\right].
\end{aligned}$$

The first two terms are clearly $O_p\left(N^{-\frac{1}{2}}\right)$. For the third we have that

$$\Pi'\Pi = \begin{pmatrix} \sum_{i=1}^{N} 1 & \sum_{i=1}^{N} \tilde{v}_i & \sum_{i=1}^{N} \tilde{v}_i^2 & \cdots & \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m} \\ \sum_{i=1}^{N} \tilde{v}_i & \sum_{i=1}^{N} \tilde{v}_i^2 & \sum_{i=1}^{N} \tilde{v}_i^3 & \cdots & \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m+1} \\ \sum_{i=1}^{N} \tilde{v}_i^2 & \sum_{i=1}^{N} \tilde{v}_i^3 & \sum_{i=1}^{N} \tilde{v}_i^4 & \cdots & \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m} & \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m+1} & \sum_{i=1}^{N} \tilde{v}_i^{\kappa_m+2} & \cdots & \sum_{i=1}^{N} \tilde{v}_i^{2\kappa_m} \end{pmatrix},$$

and its inverse can be represented as

$$(\Pi'\Pi)^{-1} = D^{-1} \begin{pmatrix} M_{11} & M_{12} & M_{13} & \cdots & M_{1\kappa_m} \\ M_{12} & M_{22} & M_{23} & \cdots & M_{2\kappa_m} \\ M_{13} & M_{23} & M_{33} & \cdots & M_{3\kappa_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{1\kappa_m} & M_{2\kappa_m} & M_{3\kappa_m} & \cdots & M_{\kappa_m\kappa_m} \end{pmatrix},$$

where D (the determinant of $\Pi'\Pi$) is the sum of $\kappa_m! N^{\kappa_m}$ monomials $\tilde{v}_{i_1}^{j_1}\tilde{v}_{i_2}^{j_2}\ldots\tilde{v}_{i_{\kappa_m}}^{j_{\kappa_m}}$, and M_{kl}(the $(k,l)^{th}$minor of $\Pi'\Pi$) is the sum of $(\kappa_m-1)!N^{\kappa_m-1}$ monomials $\tilde{v}_{i_1}^{j_1}\tilde{v}_{i_2}^{j_2}\ldots\tilde{v}_{i_{\kappa_m-1}}^{j_{\kappa_m-1}}$. As a consequence the $(k,l)^{th}$ element of $N\left\{(\Pi'\Pi)^{-1} - \left(\hat{\Pi}'\hat{\Pi}\right)^{-1}\right\}$ can be written as

$$\begin{aligned} N\frac{\left(\hat{D}-D\right)M_{kl} - D\left(\hat{M}_{kl}-M_{kl}\right)}{D\hat{D}} &= N\frac{\kappa_m!\,(\kappa_m-1)!N^{2\kappa_m-1}O_p\left(\kappa_m N^{-\frac{1}{2}}\right)}{(\kappa_m!)^2 N^{2\kappa_m}} \\ &= O_p\left(N^{-\frac{1}{2}}\right) \end{aligned}$$

since $\tilde{v}_{i_1}^{j_1}\tilde{v}_{i_2}^{j_2}\ldots\tilde{v}_{i_{\kappa_m}}^{j_{\kappa_m}} - \hat{\tilde{v}}_{i_1}^{j_1}\hat{\tilde{v}}_{i_2}^{j_2}\ldots\hat{\tilde{v}}_{i_{\kappa_m}}^{j_{\kappa_m}}$ can be decomposed in κ_m terms where only one factor is a difference, like $\left(\tilde{v}_{i_1}^{j_1} - \hat{\tilde{v}}_{i_1}^{j_1}\right)\tilde{v}_{i_2}^{j_2}\ldots\tilde{v}_{i_{\kappa_m}}^{j_{\kappa_m}}$. This proves the lemma. □

Lemma B.3. *Under the assumptions of Theorem 3.1 it holds that*

$$N^{-1}\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right) = \bar{A}_N + o_p(1).$$

Proof. First it holds that

$$N^{-1}\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right)-N^{-1}\hat{\tilde{S}}\left(\beta\right)'\hat{\tilde{S}}\left(\beta\right) \;=\; o_p\left(N^{-\frac{1}{2}}\right), \tag{B.1}$$

from Lemma B.2 and Assumption C. We further have that

$$\begin{aligned} N^{-1}\hat{\tilde{S}}\left(\beta\right)'\hat{\tilde{S}}\left(\beta\right)-N^{-1}\tilde{S}'\tilde{S} &= N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}H^{*}+N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}\tilde{S} \\ &\quad +N^{-1}\tilde{S}'\mathcal{Q}_{\Psi}H^{*}-N^{-1}\tilde{S}'P_{\Psi}\tilde{S} \\ &= o_p\left(1\right). \end{aligned} \tag{B.2}$$

This will be shown term by term for the diagonal elements. For the other elements, the result follows by the Cauchy-Schwarz inequality. For the first term it holds that

$$\begin{aligned} N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}H^{*} &= N^{-1}\begin{pmatrix} H_x'-\eta_x'\Phi_{N\kappa_\varepsilon}'\left(\tilde{v}_i\right) \\ H_z^{*\prime}-\eta_z'\Phi_{N\kappa_u}^{*\prime}\left(\tilde{v}_i\right)\end{pmatrix}\mathcal{Q}_{\Psi}\begin{pmatrix} H_x-\Phi_{N\kappa_\varepsilon}\left(\tilde{v}_i\right)\eta_x \\ H_z^{*}-\Phi_{N\kappa_u}^{*}\left(\tilde{v}_i\right)\eta_z\end{pmatrix} \\ &\leq N^{-1}\begin{pmatrix} H_x'-\eta_x'\Phi_{N\kappa_\varepsilon}'\left(\tilde{v}_i\right) \\ H_z^{*\prime}-\eta_z'\Phi_{N\kappa_u}^{*\prime}\left(\tilde{v}_i\right)\end{pmatrix}\begin{pmatrix} H_x-\Phi_{N\kappa_\varepsilon}\left(\tilde{v}_i\right)\eta_x \\ H_z^{*}-\Phi_{N\kappa_u}^{*}\left(\tilde{v}_i\right)\eta_z\end{pmatrix}. \end{aligned}$$

For its j-th diagonal element, $j\leq K_1$, we have that

$$N^{-1}\left(H_{x;j}'-\eta_{x;j}'\Phi_{N\kappa_\varepsilon}'\left(\tilde{v}_i\right)\right)\left(H_{x;j}-\Phi_{N\kappa_\varepsilon}\left(\tilde{v}_i\right)\eta_{x;j}\right) \;=\; O_p\left(e_{h_{x;j}}^2\left(\kappa_\varepsilon\left(N\right)\right)\right),$$

and for its K_1+k-th diagonal element

$$N^{-1}\left(H_{z;k}^{*\prime}-\eta_{z;k}'\Phi_{N\kappa_u}^{*\prime}\left(\tilde{v}_i\right)\right)\left(H_{z;k}^{*}-\Phi_{N\kappa_u}^{*}\left(\tilde{v}_i\right)\eta_{z;k}\right) \;=\; O_p\left(e_{h_{z;k}}^2\left(\kappa_u\left(N\right)\right)\right),$$

by the assumptions of Theorem 3.1. For the first term we thus have that

$$N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}H^{*} \;=\; o_p\left(1\right).$$

The second term is equal to

$$N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}\tilde{S} \;=\; N^{-1}\begin{pmatrix} H_x'-\eta_x'\Phi_{N\kappa_\varepsilon}'\left(\tilde{v}_i\right) \\ H_z^{*\prime}-\eta_z'\Phi_{N\kappa_u}^{*\prime}\left(\tilde{v}_i\right)\end{pmatrix}\tilde{S}.$$

The variance of its j-th element, $0 < j \leq K_1$, is given by

$$N^{-2}\mathrm{E}\left[\left(H'_{x;j} - \eta'_{x;j}\Phi'_{N\kappa_\varepsilon}\left(\tilde{v}_i\right)\right)\mathrm{E}\left[\tilde{S}\tilde{S}'\middle|\tilde{V}\right]\left(H_{x;j} - \Phi_{N\kappa_\varepsilon}\left(\tilde{v}_i\right)\eta_{x;j}\right)\right] \;\leq\; N^{-1}\Delta\left(e^2_{h_{x;j}}\left(\kappa_\varepsilon\left(N\right)\right)\right),$$

and that of its $K_1 + k$-th element, $0 < k \leq K_1 + K_2 + K_3$, by

$$N^{-2}\mathrm{E}\left[\left(H^{*\prime}_{z;k} - \eta'_{z;k}\Phi^*_{N\kappa_u}\left(\tilde{v}_i\right)\right)\mathrm{E}\left[\tilde{S}\tilde{S}'\middle|\tilde{V}\right]\left(H^*_{z;k} - \Phi^{*\prime}_{N\kappa_u}\left(\tilde{v}_i\right)\eta_{z;k}\right)\right] \;\leq\; N^{-1}\Delta\left(e^2_{h_{z;k}}\left(\kappa_u\left(N\right)\right)\right),$$

using independence and Assumption E.4 and the assumptions of Theorem 3.1. By the Cauchy-Schwarz inequality we have that the (j_1, j_2)-th element of the covariance matrix of $N^{-1}H^{*\prime}\mathcal{Q}_\Psi\tilde{S}$ is smaller than $N^{-1}\Delta\left(e_{h_{x;j_1}}\left(\kappa_\varepsilon\left(N\right)\right)e_{h_{x;j_\circ 2}}\left(\kappa_\varepsilon\left(N\right)\right)\right)$ for $j_1, j_2 \leq K_1$. The $(K_1 + k_1, K_1 + k_2)$-th element is smaller than $N^{-1}\Delta\left(e_{h_{z;k_1}}\left(\kappa_u\left(N\right)\right)e_{h_{z;k_2}}\left(\kappa_u\left(N\right)\right)\right)$. The $(j_1, K_1 + k_2)$-th element, finally, is smaller than $N^{-1}\Delta\left(e_{h_{x;j_1}}\left(\kappa_\varepsilon\left(N\right)\right)e_{h_{z;k_2}}\left(\kappa_u\left(N\right)\right)\right)$. By the Chebyshev inequality, it holds that $N^{-1}H^{*\prime}\mathcal{Q}_\Psi\tilde{S} = o_p\left(1\right)$. The same applies for the third term $\left(N^{-1}\tilde{S}'\mathcal{Q}_\Psi H^*\right)$. For the fourth term it holds that

$$\begin{aligned}
\mathrm{E}\left[N^{-1}\tilde{S}'_j P_\Psi \tilde{S}_j\right] &= N^{-1}\mathrm{E}\left[\mathrm{Tr}\left(P_\Psi \mathrm{E}\left[\tilde{S}_j\tilde{S}'_j\middle|\tilde{V}\right]P_\Psi\right)\right] \\
&\leq N^{-1}\Delta\mathrm{E}\left[\mathrm{Tr}\left(P_\Psi I_N P_\Psi\right)\right] \\
&= N^{-1}\Delta\left(\kappa_\varepsilon\left(N\right) + \kappa_u\left(N\right)\right) \\
&= o_p\left(1\right),
\end{aligned}$$

where the inequality holds by the assumptions of Theorem 3.1. Finally, by Asumption E.2 it holds that

$$N^{-1}\tilde{S}'\tilde{S} - \bar{A}_N \;=\; o_p\left(1\right). \tag{B.3}$$

Equalities (B.1), (B.2) and (B.3) entail Lemma B.3. □

B.2 Proofs of Theorems

Proof of Theorem 3.1. First, the estimator for the 'parametric part' of the parameter vector is given by (3.15) as

$$\begin{pmatrix} \hat{\alpha}_1\left(\hat{\beta}\right) - \alpha_1 \\ \hat{\gamma}\left(\hat{\beta}\right) - \gamma \end{pmatrix} = \left(\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right)\right)^{-1}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left[\left(G^* - \hat{G}^*\left(\hat{\beta}\right)\right) + \left(\Xi - \hat{\Xi}\left(\hat{\beta}\right)\right)\right].$$

The theorem will be proved by computing term by term the orders of convergence.

1. For the first factor, it holds that $\left(N^{-1}\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right)\right)^{-1} = O_p(1)$ by Lemma B.3 and Assumption E.2.

2. For the second factor we have that

$$\begin{aligned} N^{-1}S'\mathcal{Q}_{\hat{\Psi}}\left(G^* + \Xi\right) &= N^{-1}S'\mathcal{Q}_{\Psi}\left(G^* + \Xi\right) + N^{-1}S'\left(\mathcal{Q}_{\hat{\Psi}} - \mathcal{Q}_{\Psi}\right)\left(G^* + \Xi\right) \\ &= N^{-1}\left(\tilde{S} + H^*\right)'\mathcal{Q}_{\Psi}\left(G^* + \Xi\right) \\ &\quad + N^{-1}S'\left(\mathcal{Q}_{\hat{\Psi}} - \mathcal{Q}_{\Psi}\right)\left(G^* + \Xi\right) \\ &= N^{-1}\tilde{S}'\Xi - N^{-1}\tilde{S}'\mathcal{P}_{\Psi}\Xi + N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}G^* \\ &\quad + N^{-1}H^{*\prime}\mathcal{Q}_{\Psi}\Xi + N^{-1}\tilde{S}'\mathcal{Q}_{\Psi}G^* \\ &\quad + N^{-1}S'\left(\mathcal{Q}_{\hat{\Psi}} - \mathcal{Q}_{\Psi}\right)\left(G^* + \Xi\right), \end{aligned}$$

   for which the orders of magnitude will be derived term by term.

   (a) For the first term $\left(N^{-1}\tilde{S}'\Xi\right)$, we have, by Chebyshev's inequality, that

$$\Pr\left[\frac{\tilde{S}_j'\Xi}{\sqrt{N}} > \Delta\right] \leq \frac{\sum_{i=1}^{N}\mathrm{E}\left[\tilde{s}_{ji}^2\xi_i^2\right]}{N\Delta^2}.$$

   Together with Assumption E.5, this implies $\frac{\tilde{S}_j'\Xi}{\sqrt{N}} = O_p(1)$, and thus $N^{-1}\tilde{S}'\Xi = O_p\left(N^{-\frac{1}{2}}\right)$.

   (b) For the second term it holds that $\mathrm{E}\left[N^{-1}\tilde{S}'\mathcal{P}_{\Psi}\Xi\,|R\right] = 0$ and we have for every

component j of $\tilde{S}$ that

$$\begin{aligned}
\mathrm{E}\left[N^{-2}\tilde{S}_j'\mathcal{P}_\Psi \mathrm{E}\left[\Xi\Xi'\left|R\right.\right]\mathcal{P}_\Psi'\tilde{S}_j\right] &= \mathrm{E}\left[N^{-2}\tilde{S}_j'\mathcal{P}_\Psi \mathrm{Var}\left[Y\left|R\right.\right]\mathcal{P}_\Psi'\tilde{S}_j\right] \\
&\leq \Delta \mathrm{E}\left[N^{-2}\tilde{S}_j'\mathcal{P}_\Psi\tilde{S}_j\right] \\
&= \Delta \mathrm{E}\left[\mathrm{Tr}\left(N^{-2}\mathcal{P}_\Psi \mathrm{E}\left[\tilde{S}_j\tilde{S}_j'\middle|\tilde{V}\right]\right)\right] \\
&\leq \Delta^2\frac{(\kappa_\varepsilon+\kappa_u)}{N^2},
\end{aligned}$$

where the first inequality holds by Assumption E.1, and the second by Assumption E.6. Remark that $\kappa_\varepsilon+\kappa_u$ is equal to the trace of $\mathcal{P}_\Psi$. By Chebyshev's inequality, above inequalities imply $N^{-1}\tilde{S}'\mathcal{P}_\Psi\Xi = O_p\left(N^{-1}\left(\kappa_\varepsilon+\kappa_u\right)^{\frac{1}{2}}\right)$.

(c) For the third term,

$$N^{-1}H^{*\prime}\mathcal{Q}_\Psi\left(G_\varepsilon+G_u^*\right) = N^{-1}\left(H_x,H_z^*\right)'\mathcal{Q}_\Psi\left(G_\varepsilon+G_u^*\right),$$

we will show that each element of $N^{-1}H_x'\mathcal{Q}_\Psi\left(G_\varepsilon+G_u^*\right)$ is $O_p\left(e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)\left\{e_{g_\varepsilon}\left(\kappa_\varepsilon\left(N\right)\right)+e_{g_u}\left(\kappa_u\left(N\right)\right)\right\}\right)$. For H_z^* the proof is the same. It holds, for the j_{th} column in H_x, that

$$\begin{aligned}
N^{-1}H_{xj}'\mathcal{Q}_\Psi\left(G_\varepsilon+G_u^*\right) &= N^{-1}\left(H_{xj}-\Phi_{N\kappa}\eta_{xj}\right)' \\
&\quad \times\mathcal{Q}_\Psi\left(\left(G_\varepsilon-\Phi_{N\kappa_\varepsilon}\pi_\varepsilon\right)+\left(G_u^*-\Phi_{N\kappa_u}^*\pi_u\right)\right) \\
&\leq N^{-1}\left(H_{xj}-\Phi_{N\kappa}\eta_{xj}\right)' \\
&\quad \times\left(\left(G_\varepsilon-\Phi_{N\kappa_\varepsilon}\pi_\varepsilon\right)+\left(G_u^*-\Phi_{N\kappa_u}^*\pi_u\right)\right) \\
&= O_p\left(e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)e_{g_\varepsilon}\left(\kappa_\varepsilon\left(N\right)\right)\right) \\
&\quad +O_p\left(e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)e_{g_u}\left(\kappa_u\left(N\right)\right)\right),
\end{aligned}$$

since it holds that

$$\begin{aligned}
&\mathrm{E}\left[N^{-1}\left(H_{xj}-\Phi_{N\kappa}\eta_{xj}\right)'\left(G_\varepsilon-\Phi_{N\kappa}\pi_\varepsilon\right)\right] \\
\leq\ &\sqrt{\mathrm{E}\left[N^{-1}\left(H_x-\Phi_{N\kappa}\eta_{xj}\right)'\left(H_x-\Phi_{N\kappa}\eta_{xj}\right)\right]} \\
&\times\sqrt{\mathrm{E}\left[N^{-1}\left(G_\varepsilon-\Phi_{N\kappa}\pi_\varepsilon\right)'\left(G_\varepsilon-\Phi_{N\kappa}\pi_\varepsilon\right)\right]} \\
\leq\ &e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)e_{g_\varepsilon}\left(\kappa_\varepsilon\left(N\right)\right),
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second from Assumption E. It thus holds that

$$\begin{aligned} & N^{-1}H_z^{*\prime}\mathcal{Q}_\Psi\left(G_\varepsilon + G_u^*\right) \\ = \; & O_p\left(\left\{e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right) + e_{h_z}\left(\kappa_u\left(N\right)\right)\right\}\left\{e_{g_\varepsilon}\left(\kappa_\varepsilon\left(N\right)\right) + e_{g_u}\left(\kappa_u\left(N\right)\right)\right\}\right). \end{aligned}$$

(d) For the fourth term, $N^{-1}H^{*\prime}\mathcal{Q}_\Psi\Xi = N^{-1}\left(H_\varepsilon, H_u^*\right)'\mathcal{Q}_\Psi\Xi$, it holds that

$$\mathrm{E}\left[N^{-1}H^{*\prime}\mathcal{Q}_\Psi\Xi\,|R\right] \;=\; 0.$$

Because the first K_1 components of this term can be written as

$$N^{-1}H_{xj}'\mathcal{Q}_\Psi\Xi \;=\; N^{-1}\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)'\mathcal{Q}_\Psi\Xi,$$

for which it holds that

$$\begin{aligned} & \mathrm{E}\left[N^{-2}\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)'\mathcal{Q}_\Psi\mathrm{E}\left[\Xi\Xi'\,|R\right]\mathcal{Q}_\Psi\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)\right] \\ \leq \; & \Delta\mathrm{E}\left[N^{-2}\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)'\mathcal{Q}_\Psi\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)\right] \\ \leq \; & \Delta\mathrm{E}\left[N^{-2}\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)'\left(H_{xj} - \Phi_{N\kappa}\eta_{xj}\right)\right] \\ \leq \; & \Delta\frac{e_{h_x}^2\left(\kappa_\varepsilon\left(N\right)\right)}{N}, \end{aligned}$$

and, by Markov's inequality, that

$$N^{-1}H_{xj}'\mathcal{Q}_\Psi\Xi \;=\; O_p\left(N^{-\frac{1}{2}}e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)\right).$$

Similarly, for the last $K_1 + K_2 + K_3$ components we have that

$$\mathrm{E}\left[N^{-2}\left(H_{zl} - \Phi_{N\kappa}^*\eta_{zl}\right)'\mathcal{Q}_\Psi\mathrm{E}\left[\Xi\Xi'\,|R\right]\mathcal{Q}_\Psi'\left(H_{zl} - \Phi_{N\kappa}^*\eta_{zl}\right)\right] \;\leq\; \Delta\frac{e_{h_z}^2\left(\kappa_u\left(N\right)\right)}{N}.$$

As a result it holds that

$$N^{-1}H^{*\prime}\mathcal{Q}_\Psi\Xi \;=\; O_p\left(N^{-\frac{1}{2}}e_{h_x}\left(\kappa_\varepsilon\left(N\right)\right)\right) + O_p\left(N^{-\frac{1}{2}}e_{h_z}\left(\kappa_u\left(N\right)\right)\right).$$

(e) The fifth term is $O_p\left(N^{-\frac{1}{2}}e_{g_\varepsilon}\left(\kappa_\varepsilon\left(N\right)\right)\right) + O_p\left(N^{-\frac{1}{2}}e_{g_u}\left(\kappa_u\left(N\right)\right)\right)$ by a similar argument.

(f) The last term can be written as

$$\begin{aligned} N^{-1}S'\left(\mathcal{Q}_{\hat{\Psi}}-\mathcal{Q}_{\Psi}\right)\left(G_{\varepsilon}+G_{u}^{*}+\Xi\right) &= N^{-1}S'\left(\mathcal{P}_{\Psi}-\mathcal{P}_{\hat{\Psi}}\right)\left(G_{\varepsilon}+G_{u}^{*}+\Xi\right) \\ &= O_{p}\left(N^{-\frac{1}{2}}\right), \end{aligned}$$

by Lemma B.2.

As a final result we have that

$$\begin{aligned} \begin{pmatrix} \hat{\alpha}_{1}\left(\hat{\beta}\right)-\alpha_{1} \\ \hat{\gamma}\left(\hat{\beta}\right)-\gamma \end{pmatrix} &= O_{p}\left(N^{-\frac{1}{2}}\right)+O_{p}\left(N^{-1}\left(\kappa_{\varepsilon}+\kappa_{u}\right)^{\frac{1}{2}}\right) \\ &\quad +O_{p}\left(\left\{e_{h_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)+e_{h_{u}}\left(\kappa_{u}\left(N\right)\right)\right\}\left\{e_{g_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)+e_{g_{u}}\left(\kappa_{u}\left(N\right)\right)\right\}\right) \\ &\quad +O_{p}\left(N^{-\frac{1}{2}}e_{h_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\right)+O_{p}\left(N^{-\frac{1}{2}}e_{h_{u}}\left(\kappa_{u}\left(N\right)\right)\right) \\ &\quad +O_{p}\left(N^{-\frac{1}{2}}e_{g_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\right)+O_{p}\left(N^{-\frac{1}{2}}e_{g_{u}}\left(\kappa_{u}\left(N\right)\right)\right). \qquad \square \end{aligned}$$

In order to prove Theorem 3.2, we first prove the following lemma.

Lemma B.4. *It holds that*

$$N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(G^{*}-\hat{G}^{*}\left(\hat{\beta}\right)\right) = D\left[\sqrt{N}\left(\hat{\beta}-\beta\right)\right]+o_{p}\left(1\right), \tag{B.4}$$

$$\sqrt{N}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(\Xi-\hat{\Xi}\right) = N^{-\frac{1}{2}}\tilde{S}'\Xi+o_{p}\left(1\right), \tag{B.5}$$

$$N^{-\frac{1}{2}}\tilde{S}'\Xi+D\left[\sqrt{N}\left(\hat{\beta}-\beta\right)\right] \xrightarrow{d} N\left(0,\bar{B}_{N}+D\Sigma_{\beta\beta}D'\right). \tag{B.6}$$

Proof.

Proof of (B.4). It holds that the left-hand side of (B.4) equals

$$\begin{aligned} N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(G^{*}-\hat{G}^{*}\left(\hat{\beta}\right)\right) &= N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\beta\right)'\left(G^{*}-\hat{G}^{*}\left(\beta\right)\right) \\ &\quad +N^{-\frac{1}{2}}\left(\hat{H}^{*}\left(\beta\right)-\hat{H}^{*}\left(\hat{\beta}\right)\right)'\left(G^{*}-\hat{G}^{*}\left(\beta\right)\right) \\ &\quad +N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(\hat{G}^{*}\left(\beta\right)-\hat{G}^{*}\left(\hat{\beta}\right)\right). \end{aligned}$$

For the first term we have that

$$\begin{aligned} N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\beta\right)'\left(G^{*}-\hat{G}^{*}\left(\beta\right)\right) &= N^{-\frac{1}{2}}\tilde{S}'\mathcal{Q}_{\Psi}G^{*}+N^{-\frac{1}{2}}H^{*\prime}\mathcal{Q}_{\Psi}G^{*} \\ &\leq O_{p}\left(e_{h_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)e_{g_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\sqrt{N}\right) \\ &\quad +O_{p}\left(e_{h_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)e_{g_{u}}\left(\kappa_{u}\left(N\right)\right)\sqrt{N}\right) \\ &\quad +O_{p}\left(e_{h_{u}}\left(\kappa_{u}\left(N\right)\right)e_{g_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\sqrt{N}\right) \\ &\quad +O_{p}\left(e_{h_{u}}\left(\kappa_{u}\left(N\right)\right)e_{g_{u}}\left(\kappa_{u}\left(N\right)\right)\sqrt{N}\right) \\ &\quad +O_{p}\left(e_{g_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\right)+O_{p}\left(e_{g_{u}}\left(\kappa_{u}\left(N\right)\right)\right), \end{aligned}$$

from the proof of Theorem 3.1, terms 3 and 5. For the second term it holds that

$$\begin{aligned} N^{-\frac{1}{2}}\left(\hat{H}^{*}\left(\beta\right)-\hat{H}^{*}\left(\hat{\beta}\right)\right)'\left(G^{*}-\hat{G}^{*}\left(\beta\right)\right) &= N^{-\frac{1}{2}}H^{*\prime}\left(\mathcal{P}_{\Psi}-\mathcal{P}_{\hat{\Psi}}\right)\left(G^{*}-\hat{G}^{*}\left(\beta\right)\right) \\ &= o_{p}\left(1\right), \end{aligned}$$

by the Cauchy-Schwarz inequality and Assumption D. Defining

$$D = \mathrm{E}\left[\tilde{s}_{i}\left.\frac{\partial g^{*}\left(x\right)}{\partial x}\right|_{x=\tilde{v}_{i}}\frac{\partial\tilde{v}_{i}}{\partial\beta'}\right],$$

we have that,

$$N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(\hat{G}^{*}\left(\beta\right)-\hat{G}^{*}\left(\hat{\beta}\right)\right) = \sqrt{N}D\left(\hat{\beta}-\beta\right)+o_{p}\left(1\right). \qquad \square$$

Proof of (B.5). It holds that

$$\begin{aligned} N^{-\frac{1}{2}}\hat{\tilde{S}}\left(\hat{\beta}\right)'\left(\Xi-\hat{\Xi}\right) &= N^{-\frac{1}{2}}S'\mathcal{Q}_{\hat{\Psi}}\Xi-N^{-\frac{1}{2}}S'\mathcal{Q}_{\hat{\Psi}}\mathcal{P}_{\hat{\Psi}}\Xi \\ &= N^{-\frac{1}{2}}S'\mathcal{Q}_{\Psi}\Xi+N^{-\frac{1}{2}}S'\left(\mathcal{Q}_{\hat{\Psi}}-\mathcal{Q}_{\Psi}\right)\Xi-N^{-\frac{1}{2}}S'\mathcal{Q}_{\hat{\Psi}}\mathcal{P}_{\hat{\Psi}}\Xi, \end{aligned}$$

where the second term $N^{-\frac{1}{2}}S'\left(\mathcal{Q}_{\hat{\Psi}}-\mathcal{Q}_{\Psi}\right)\Xi=o_{p}\left(1\right)$, by Lemma B.2, and the third term $N^{-\frac{1}{2}}S'\mathcal{Q}_{\hat{\Psi}}\mathcal{P}_{\hat{\Psi}}\Xi=o_{p}\left(1\right)$ by virtue of the equality $\mathcal{P}_{\hat{\Psi}}\mathcal{Q}_{\hat{\Psi}}=0$. The first term can be rewritten as

$$\begin{aligned} N^{-\frac{1}{2}}S'\mathcal{Q}_{\Psi}\Xi &= N^{-\frac{1}{2}}\tilde{S}'\Xi-N^{-\frac{1}{2}}\tilde{S}'\mathcal{P}_{\Psi}\Xi+N^{-\frac{1}{2}}H^{*\prime}\mathcal{Q}_{\Psi}\Xi \\ &= N^{-\frac{1}{2}}\tilde{S}'\Xi+o_{p}\left(1\right)+O_{p}\left(e_{h_{\varepsilon}}\left(\kappa_{\varepsilon}\left(N\right)\right)\sqrt{N}\right)+O_{p}\left(e_{h_{u}}\left(\kappa_{u}\left(N\right)\right)\sqrt{N}\right). \qquad \square \end{aligned}$$

Proof of (B.6). This expression holds by (B.4)-(B.5), by Assumption E.3, F and by the asymptotic independence of both terms. □

Proof of Theorem 3.2. Consider now

$$\begin{aligned}\sqrt{N}\begin{pmatrix}\hat{\alpha}_1\left(\hat{\beta}\right)-\alpha_1\\ \hat{\gamma}\left(\hat{\beta}\right)-\gamma\end{pmatrix} &= \sqrt{N}\left(\hat{\tilde{S}}\left(\hat{\beta}\right)'\hat{\tilde{S}}\left(\hat{\beta}\right)\right)^{-1}\\ &\quad\times\hat{\tilde{S}}\left(\hat{\beta}\right)'\left[\left(G^*-\hat{G}^*\left(\hat{\beta}\right)\right)+\left(\Xi-\hat{\Xi}\left(\hat{\beta}\right)\right)\right]\\ &= \left(N^{-1}\tilde{S}'\tilde{S}+o_p\left(1\right)\right)^{-1}\left\{D\left[\sqrt{N}\left(\hat{\beta}-\beta\right)\right]+N^{-\frac{1}{2}}\tilde{S}'\Xi+o_p\left(1\right)\right\}\\ &\xrightarrow{d} N\left(0,\bar{A}_N^{-1}\left(\bar{B}_N+D\Sigma_{\beta\beta}D'\right)\bar{A}_N^{-1}\right),\end{aligned}$$

by Lemma B.3 and B.4. From Assumption C, we have that

$$\mathrm{Cov}\left[\sqrt{N}\left(\hat{\alpha}_1'\left(\hat{\beta}\right)-\alpha_1',\hat{\gamma}'\left(\hat{\beta}\right)-\gamma'\right)',\sqrt{N}\left(\hat{\beta}-\beta\right)\right] = \bar{A}_N^{-1}D\mathrm{Var}\left[\sqrt{N}\left(\hat{\beta}-\beta\right)\right],$$

and Theorem 3.2 follows by the Lyapunov CLT. □

Proof of Theorem 3.3. It holds that

$$\begin{aligned}\mathcal{P}_{\hat{\Psi}}Y &= \mathcal{P}_{\hat{\Psi}}S\begin{pmatrix}\alpha_1\\ \gamma\end{pmatrix}+\mathcal{P}_{\hat{\Psi}}G^*+\mathcal{P}_{\hat{\Psi}}\Xi\\ &= \mathcal{P}_{\hat{\Psi}}S\begin{pmatrix}\hat{\alpha}_1\left(\hat{\beta}\right)\\ \hat{\gamma}\left(\hat{\beta}\right)\end{pmatrix}+\mathcal{P}_{\hat{\Psi}}\hat{G}^*.\end{aligned}$$

Rearranging terms and substraction of $\mathcal{Q}_{\hat{\Psi}}G^*$ and making use of the decomposition $S=\tilde{S}+H^*$ results in

$$\left(\hat{G}^*-G^*\right) = \mathcal{P}_{\hat{\Psi}}S\begin{pmatrix}\alpha_1-\hat{\alpha}_1\left(\hat{\beta}\right)\\ \gamma-\hat{\gamma}\left(\hat{\beta}\right)\end{pmatrix}-\mathcal{Q}_{\hat{\Psi}}G^*+\mathcal{P}_{\hat{\Psi}}\Xi.$$

We further have that

$$
\begin{aligned}
& N^{-1}\begin{pmatrix} \alpha_1 - \hat{\alpha}_1\left(\hat{\beta}\right) \\ \gamma - \hat{\gamma}\left(\hat{\beta}\right) \end{pmatrix}' S'\mathcal{P}_{\hat{\Psi}} S \begin{pmatrix} \alpha_1 - \hat{\alpha}_1\left(\hat{\beta}\right) \\ \gamma - \hat{\gamma}\left(\hat{\beta}\right) \end{pmatrix} \\
= \quad & O_p\left(N^{-1}\right) + O_p\left(K_\varepsilon(N)\, N^{-2}\right) + O_p\left(K_u(N)\, N^{-2}\right) \\
+ \quad & O_p\left(e^2_{h_\varepsilon}\left(\kappa_\varepsilon(N)\right) e^2_{g_\varepsilon}\left(\kappa_\varepsilon(N)\right)\right) + O_p\left(e^2_{h_\varepsilon}\left(\kappa_\varepsilon(N)\right) e^2_{g_u}\left(\kappa_u(N)\right)\right) \\
+ \quad & O_p\left(e^2_{h_u}\left(\kappa_u(N)\right) e^2_{g_\varepsilon}\left(\kappa_\varepsilon(N)\right)\right) + O_p\left(e^2_{h_u}\left(\kappa_u(N)\right) e^2_{g_u}\left(\kappa_u(N)\right)\right) \\
+ \quad & O_p\left(N^{-1} e^2_{h_\varepsilon}\left(\kappa_\varepsilon(N)\right)\right) + O_p\left(N^{-1} e^2_{h_u}\left(\kappa_u(N)\right)\right) \\
+ \quad & O_p\left(N^{-1} e^2_{g_\varepsilon}\left(\kappa_\varepsilon(N)\right)\right) + O_p\left(N^{-1} e^2_{g_u}\left(\kappa_u(N)\right)\right),
\end{aligned}
$$

since $N^{-1}S'\mathcal{P}_{\hat{\Psi}}S$ is bounded in probability and by Theorem 3.1. For the second term it holds that

$$
\begin{aligned}
N^{-1}G^{*\prime}\mathcal{Q}_{\hat{\Psi}}G^* &= N^{-1}\left(G'_\varepsilon \mathcal{Q}_{\hat{\Psi}} G_\varepsilon + 2G'_\varepsilon \mathcal{Q}_{\hat{\Psi}} G^*_u + G^{*\prime}_u \mathcal{Q}_{\hat{\Psi}} G^*_u\right) \\
&= O_p\left(e^2_{g_\varepsilon}\left(\kappa_\varepsilon(N)\right)\right) + O_p\left(e_{g_\varepsilon}\left(\kappa_u(N)\right) e_{g_u}\left(\kappa_u(N)\right)\right) + O_p\left(e^2_{g_u}\left(\kappa_u(N)\right)\right),
\end{aligned}
$$

and for the last term that $N^{-1}\Xi'\mathcal{P}_{\hat{\Psi}}\Xi = O_p\left(\left(K_\varepsilon + K_u\right)N^{-1}\right)$. Theorem 3.3 follows from the triangle inequality. $\square$

Appendix C

IV Estimation in Panels: Proofs of Theorems

C.1 Identification

Proof of Proposition 4.1. A necessary and sufficient condition for identifiability of φ in (4.9), is that $\mathcal{P}_V R$ is of full column rank. Given Assumption A-D, a sufficient prerequisite for this is that both $\mathcal{P}_{V_X} X$ and $\mathcal{P}_{V_S} S$ are of full column rank. Since $\mathcal{P}_i$, $\mathcal{P}_p$ and $\mathcal{P}_r$ are mutually orthogonal, variables lying in one of the three subspaces can only have instruments from that subspace and variables spanning two subspaces cannot have instruments from the third subspace. Consider now all possible spaces generated by all possible unions of these mutually orthogonal subspaces, then for each of these unions it must hold that the number of variables lying within it should not be greater than the number of instruments generated from it.

More specifically, the number of variables in X that only have a nonzero component in the individual-specific dimension, $k_{a\ddot{b}\ddot{c}} + k_{\dot{a}\ddot{b}\ddot{c}}$, should not be greater than the number of individual-specific instruments, k_a. Since we have that $k_{\ddot{a}\ddot{b}\ddot{c}} = 0$, the resulting identification restriction is given by

$$k_{\ddot{b}\ddot{c}} \quad \leq \quad k_a. \tag{C.1}$$

The same holds for S ($j_{\ddot{b}\ddot{c}} \leq j_a$) and the same condition holds for one component variables

along other dimensions

$$m_{\ddot{a}\ddot{c}} \leq m_b, \tag{C.2}$$
$$m_{\ddot{a}\ddot{b}} \leq m_c, \tag{C.3}$$

for $m = j, k$. The number of variables in X that miss the individual-specific dimension is given by $k_{\ddot{a}}$, and it should not be greater than the sum of the period-specific and the idiosyncratic instruments $(k_b + k_c)$

$$k_{\ddot{a}} \leq k_b + k_c, \tag{C.4}$$

and similarly for variables in S that miss the individual-specific dimension. Analogous conditions are needed for variables also missing exactly one other dimension

$$m_{\dot{b}} \leq m_a + m_c, \tag{C.5}$$
$$m_{\dot{c}} \leq m_a + m_b, \tag{C.6}$$

for $m = j, k$. Finally, the number of variables in X and S needs to be smaller than the number of instruments for X, respectively S,

$$\begin{aligned} K &\leq k_{V_X}, \\ J &\leq j_{V_S}. \end{aligned} \tag{C.7}$$

which, after simple algebra, leads to

$$\begin{aligned} & m_{\dot{a}\dot{b}\dot{c}} + (m_{\ddot{a}\dot{b}\dot{c}} + m_{\dot{a}\ddot{b}\dot{c}} + m_{\dot{a}\dot{b}\ddot{c}}) + (m_{\dot{a}\ddot{b}\ddot{c}} + m_{\ddot{a}\dot{b}\ddot{c}} + m_{\ddot{a}\ddot{b}\dot{c}}) \\ \leq \; & 2m_{abc} + (m_{\dot{a}bc} + m_{a\dot{b}c} + m_{ab\dot{c}}) + (m_{\ddot{a}bc} + m_{a\ddot{b}c} + m_{ab\ddot{c}}), \end{aligned}$$

for $m = j, k$.

Having established the identification of X and S separately, we can easily verify that these conditions imply the identification of

$$r_{it} = \left(1, x_{it}', (s_{it} - \mu_S)', \text{Vech}\left[x_{it} \cdot (x_{it} - \mu_X)'\right]', \text{Vec}\left[x_{it} \cdot (s_{it} - \mu_S)'\right]'\right)'$$

by

$$v_{it} \quad = \quad \left(1, v'_{X,it}, v'_{S,it}, \text{Vech}\left[v_{X,it} \cdot v'_{X,it}\right]', \text{Vec}\left[v_{X,it} \cdot v'_{S,it}\right]'\right)'.$$

Consider first variables with only one component. For instance, the number of individual-specific variables in R is given by the number of individual-specific variables in X ($k_{\ddot{b}\ddot{c}}$), plus the number of individual-specific variables in S ($j_{\ddot{b}\ddot{c}}$), plus the number of individual-specific variables in P ($k_{\ddot{b}\ddot{c}} \cdot (k_{\ddot{b}\ddot{c}} + 1)/2$), plus the number of individual-specific variables in Q ($j_{\ddot{b}\ddot{c}} \cdot k_{\ddot{b}\ddot{c}}$) and totals $(1 + k_{\ddot{b}\ddot{c}})(1 + k_{\ddot{b}\ddot{c}}/2 + j_{\ddot{b}\ddot{c}}) - 1$. Also, the number of individual-specific instruments in v_{it} is equal to the number of individual-specific instruments in $v_{X,it}$ (k_a), plus the number of individual-specific instruments in $v_{S,it}$ (j_a), plus the number of individual-specific instruments in $\text{Vech}\left[v_{X,it} \cdot v'_{X,it}\right]$ ($k_a(k_a + 1)/2$), plus the number of individual-specific instruments in $\text{Vec}\left[v_{X,it} \cdot v'_{S,it}\right]$ ($j_a \cdot k_a$) and is equal to $(1 + k_a)(1 + k_a/2 + j_a) - 1$. It is now easy to verify that the identification restriction

$$(1 + k_{\ddot{b}\ddot{c}})(1 + k_{\ddot{b}\ddot{c}}/2 + j_{\ddot{b}\ddot{c}}) - 1 \quad \leq \quad (1 + k_a)(1 + k_a/2 + j_a) - 1$$

is fulfilled when the conditions $k_{\ddot{b}\ddot{c}} \leq k_a$ and $j_{\ddot{b}\ddot{c}} \leq j_a$ are fulfilled. The reasoning is analogous for the period-specific and for the idiosyncratic variables in R.

Consider now variables with two components. For instance, the total number of variables without idiosyncratic component in R is given by the number of variables without idiosyncratic component in X ($k_{\ddot{c}}$) and in S ($j_{\ddot{c}}$), plus the number of individual-specific variables in $\text{Vech}\left[x_{it} \cdot x'_{it}\right]$ ($k_{\ddot{b}\ddot{c}}(k_{\ddot{b}\ddot{c}} + 1)/2$) and in $\text{Vec}\left[x_{it} \cdot s'_{it}\right]$ ($j_{\ddot{b}\ddot{c}} \cdot k_{\ddot{b}\ddot{c}}$), plus the number of period-specific variables in $\text{Vech}\left[x_{it} \cdot x'_{it}\right]$ ($k_{\ddot{a}\ddot{c}}(k_{\ddot{a}\ddot{c}} + 1)/2$) and in $\text{Vec}\left[x_{it} \cdot s'_{it}\right]$ ($j_{\ddot{a}\ddot{c}} \cdot k_{\ddot{a}\ddot{c}}$). It should be smaller then or equal to the number of instruments without idiosyncratic component in v_{it}, which is equal to the number of instruments without idiosyncratic component in $v_{X,it}$ ($k_a + k_b$), $v_{S,it}$ ($j_a + j_b$), $\text{Vech}\left[v_{X,it} \cdot v'_{X,it}\right]$ ($k_a(k_a + 1)/2$), $\text{Vech}\left[v_{S,it} \cdot v'_{S,it}\right]$ ($k_b(k_b + 1)/2$) and $\text{Vec}\left[v_{X,it} \cdot v'_{S,it}\right]$ ($k_a \cdot j_a + k_b \cdot j_b$). The resulting identification restriction is given by

$$\begin{aligned} & k_{\ddot{c}} + j_{\ddot{c}} + k_{\ddot{a}\ddot{c}}(k_{\ddot{a}\ddot{c}} + 1)/2 + k_{\ddot{b}\ddot{c}}(k_{\ddot{b}\ddot{c}} + 1)/2 + j_{\ddot{a}\ddot{c}} \cdot k_{\ddot{a}\ddot{c}} + j_{\ddot{b}\ddot{c}} \cdot k_{\ddot{b}\ddot{c}} \\ \leq \quad & k_a + k_b + j_a + j_b + k_a(k_a + 1)/2 + k_b(k_b + 1)/2 + k_a \cdot j_a + k_b \cdot j_b, \end{aligned}$$

which is easily verified by (C.1), (C.2) and (C.6). The resulting inequalities for the variables in R without individual-specific component or without period-specific component are also easily verified.

Finally, the total number of variables should not be greater than the total number of instruments:

$$(1+K)\left(1+\frac{K}{2}+J\right) \;\leq\; (1+k_{V_X})\left(1+\frac{k_{V_X}}{2}+k_{V_S}\right),$$

which follows trivially from (C.7), which completes the proof of Proposition 4.1. ■

C.2 Estimation

Throughout the rest of this appendix denote $\mathcal{P}_V R$ by $\hat{R}$.

C.2.1 IV Estimation

Proof of Theorem 4.1. The consistency of

$$\hat{\varphi}_{IV} \;=\; \left(R'\mathcal{P}_V R\right)^{-1} R'\mathcal{P}_V Y$$

follows immediately from Assumption B.1-B.3, C and the choice of the instruments. In order to derive the asymptotic distribution of $\hat{\varphi}_{IV}$, define

$$C \;=\; (NT)^{-1}\,\Psi_C \begin{pmatrix} \iota'_{NT}\mathcal{P}_i H \\ \iota'_{NT}\mathcal{P}_p H \\ \iota'_{NT}\mathcal{P}_r H \end{pmatrix},$$

with $\Psi_C = \mathrm{Diag}_{j=1}^{3}\left[\sqrt{N_j}\right]$, and $\Sigma_C = \mathrm{E}\left[CC'\right]$, with N_1, N_2 and N_3 to be determined shortly. From the construction of $\mathcal{P}_i$, $\mathcal{P}_p$ and $\mathcal{P}_r$ in (4.6) we have, in general, that

$$\mathcal{P}_i H = \mathcal{P}_i\left(H_\lambda + H_\xi\right) \;\Leftrightarrow\; \mathcal{P}_r H = \mathcal{P}_r\left(H_\lambda + H_\xi\right) \tag{C.8}$$

$$\mathcal{P}_p H = \mathcal{P}_p\left(H_\tau + H_\xi\right) \;\Leftrightarrow\; \mathcal{P}_r H = \mathcal{P}_r\left(H_\tau + H_\xi\right). \tag{C.9}$$

The condition $\mathcal{P}_i H = \mathcal{P}_i\left(H_\lambda + H_\xi\right)$ verifies the presence of period-specific components in the error term H. It is satisfied if the only coefficients that show period-specific variation are the ones associated with period-specific variables or with the constant term. In that case H_τ is also period-specific and thus orthogonal to $\mathcal{P}_i$ and $\mathcal{P}_r$. Similarly, $\mathcal{P}_p H = \mathcal{P}_p\left(H_\tau + H_\xi\right)$ is satisfied if the only coefficients that show individual-specific variation are the ones associated with individual-specific variables or with the constant term. Fulfillment of both

conditions implies that only the constant term has a random coefficient associated with it, i.e. the model is an error components model.
It also holds, by Assumption A-C, that

$$\begin{aligned}\lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' H_\lambda H_\lambda' \iota_{NT}\right] &= O\left(NT^2\right)\\ \lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' H_\tau H_\tau' \iota_{NT}\right] &= O\left(N^2T\right)\end{aligned}$$

$$\lim_{N,T\to\infty} \mathrm{E}\left[\iota_{NT}' H_\xi H_\xi' \iota_{NT}\right] = O\left(NT\right), \tag{C.10}$$

which is a consequence of the block-diagonal structure of $\mathrm{E}\left[H_\lambda H_\lambda'\right]$, the permuted block-diagonal structure of $\mathrm{E}\left[H_\tau H_\tau'\right]$ and the diagonal structure of $\mathrm{E}\left[H_\xi H_\xi'\right]$.

1. In the first case, we have that $\mathcal{P}_i H = \mathcal{P}_i\left(H_\lambda + H_\xi\right)$ and $\mathcal{P}_p H = \mathcal{P}_p\left(H_\tau + H_\xi\right)$. Relations (C.8)-(C.9) imply that $\mathcal{P}_r H = \mathcal{P}_r H_\xi$ and it holds that

$$C = \left(NT\right)^{-1} \Psi_C \begin{pmatrix} \iota_{NT}' \mathcal{P}_i \left(H_\lambda + H_\xi\right) \\ \iota_{NT}' \mathcal{P}_p \left(H_\tau + H_\xi\right) \\ \iota_{NT}' \mathcal{P}_r H_\xi \end{pmatrix}.$$

   Consequently,

$$\lim_{N,T\to\infty}\left[\Sigma_C\right] = \Psi_C \begin{pmatrix} O\left(N^{-1}\right) & O\left(\left(NT\right)^{-1}\right) & O\left(\left(NT\right)^{-1}\right) \\ O\left(\left(NT\right)^{-1}\right) & O\left(T^{-1}\right) & O\left(\left(NT\right)^{-1}\right) \\ O\left(\left(NT\right)^{-1}\right) & O\left(\left(NT\right)^{-1}\right) & O\left(\left(NT\right)^{-1}\right) \end{pmatrix} \Psi_C.$$

   Now, since $\hat{R} = O_p\left(1\right)$, and by Assumption B.1-B.3, C, entries in

$$\Gamma = \lim_{N,T\to\infty} \mathrm{E}\left[\left(NT\right)^{-2} \Psi R' \mathcal{P}_V \Omega_\eta \mathcal{P}_V R \Psi\right]$$

   corresponding to individual-specific variables will be of the same order as the $(1,1)$-th element of $\lim_{N,T\to\infty}\left[\Sigma_C\right]$, period-specific variables correspond to the $(2,2)$-th element and idiosyncratic variables to the $(3,3)$-th element. In order that $\Gamma = O\left(1\right)$, we need to impose that $N_1 = N$, $N_2 = T$ and $N_3 = NT$. Variables with more than one component belong to the group of variables for which N_i is the lowest. For instance, a variable with both an individual-specific and a period-specific component will be

part of R_i if $N = o(T)$, since insuch case it holds that $T^{-1} = o(N^{-1})$. Such a variable will be an element of R_p if $T = o(N)$. Consequently, we have that $R_i = \left(R^a, R^{\mathring{a}}\right)$, $R_p = \left(R^{\ddot{a}b}, R^{\ddot{a}\dot{b}}\right)$, for $N = o(T)$ and $R_i = \left(R^{\ddot{b}a}, R^{\ddot{b}\mathring{a}}\right)$, $R_p = \left(R^b, R^{\dot{b}}\right)$, if $T = o(N)$ and always that $R_r = R^{\ddot{a}\ddot{b}}$.

The expression for Γ in the first case can be simplified, by noticing that

$$R'\mathcal{P}_V H = R'\mathcal{P}_{V_i}(\Lambda + \Xi) + R'\mathcal{P}_{V_p}(\Upsilon + \Xi) + R'\mathcal{P}_{V_r}\Xi.$$

Consequently, it holds that

$$\begin{aligned}
\Gamma &= \lim_{N,T\to\infty} \mathrm{E}\left[(NT)^{-2}\Psi R'\mathcal{P}_V\Omega_\eta\mathcal{P}_V R\Psi\right] \\
&= \lim_{N,T\to\infty} \mathrm{E}\left[\left(T\sigma_\lambda^2 + \sigma_\xi^2\right)(NT)^{-2}\Psi R'\mathcal{P}_{V_i}\mathcal{P}_i\mathcal{P}_{V_i}R\Psi\right] \\
&\quad + \lim_{N,T\to\infty} \mathrm{E}\left[\left(N\sigma_\tau^2 + \sigma_\xi^2\right)(NT)^{-2}\Psi R'\mathcal{P}_{V_p}\mathcal{P}_p\mathcal{P}_{V_p}R\Psi\right] \\
&\quad + \lim_{N,T\to\infty} \mathrm{E}\left[\sigma_\xi^2(NT)^{-2}\Psi R'\mathcal{P}_{V_r}\mathcal{P}_r\mathcal{P}_{V_r}R\Psi\right]
\end{aligned}$$

and further simplification results in $\Gamma = \mathrm{Diag}_{j=1}^{3}[M_j]$, with $M_1 = \sigma_\lambda^2 P_i$, $M_2 = \sigma_\tau^2 P_p$, $M_3 = \sigma_\xi^2 P_r$ where $P_j = \mathrm{plim}_{N,T\to\infty}\left[(NT)^{-1} R_j'\mathcal{P}_{V_j}R_j\right]$ for $j = i, p, r$.

2. In the second case we have that $\mathcal{P}_i H = \mathcal{P}_i(H_\lambda + H_\xi)$, but $\mathcal{P}_p H \neq \mathcal{P}_p(H_\tau + H_\xi)$, resulting in

$$C = (NT)^{-1}\Psi_C \begin{pmatrix} \iota_{NT}'\mathcal{P}_i(H_\lambda + H_\xi) \\ \iota_{NT}'\mathcal{P}_p(H_\lambda + H_\tau + H_\xi) \\ \iota_{NT}'\mathcal{P}_r(H_\lambda + H_\xi) \end{pmatrix}.$$

By the same reasoning as in case 1, it holds we have here that

$$\lim_{N,T\to\infty}[\Sigma_C] = \Psi_C \begin{pmatrix} O(N^{-1}) & O(N^{-1}) & O(N^{-1}) \\ O(N^{-1}) & O(N^{-1}+T^{-1}) & O(N^{-1}) \\ O(N^{-1}) & O(N^{-1}) & O(N^{-1}) \end{pmatrix}\Psi_C,$$

$N_1 = N_3 = N$, $N_2 = \min(N, T)$ and $R_p = \left(R^b, R^{\dot{b}}\right)$, if $T = o(N)$.

3. In the third case we have that $\mathcal{P}_p H = \mathcal{P}_p(H_\tau + H_\xi)$, but $\mathcal{P}_i H \neq \mathcal{P}_i(H_\lambda + H_\xi)$, which

gives

$$C = (NT)^{-1} \Psi_C \begin{pmatrix} \iota_{NT}' \mathcal{P}_i (H_\lambda + H_\tau + H_\xi) \\ \iota_{NT}' \mathcal{P}_p (H_\tau + H_\xi) \\ \iota_{NT}' \mathcal{P}_r (H_\tau + H_\xi) \end{pmatrix}.$$

Consequently, it holds that

$$\lim_{N,T\to\infty} [\Sigma_C] = \Psi_C \begin{pmatrix} O(N^{-1} + T^{-1}) & O(T^{-1}) & O(T^{-1}) \\ O(T^{-1}) & O(T^{-1}) & O(T^{-1}) \\ O(T^{-1}) & O(T^{-1}) & O(T^{-1}) \end{pmatrix} \Psi_C,$$

with $N_1 = \min(N,T)$, $N_2 = N_3 = T$ and $R_i = \left(R^a, R^{\dot{a}}\right)$, if $N = o(T)$.

4. In the fourth case, finally, it holds that $\mathcal{P}_i H \neq \mathcal{P}_i (H_\lambda + H_\xi)$ and $\mathcal{P}_p H \neq \mathcal{P}_p (H_\tau + H_\xi)$. Therefore, we have that

$$C = (NT)^{-1} \Psi_C \begin{pmatrix} \iota_{NT}' \mathcal{P}_i (H_\lambda + H_\tau + H_\xi) \\ \iota_{NT}' \mathcal{P}_p (H_\lambda + H_\tau + H_\xi) \\ \iota_{NT}' \mathcal{P}_r (H_\lambda + H_\tau + H_\xi) \end{pmatrix}$$

$\lim_{N,T\to\infty} [\Sigma_C] = O(N^{-1} + T^{-1}) \Psi_C J_3 \Psi_C$, with $J_3 = \iota_3 \iota_3'$ and thus $N_j = \min(N,T)$ for $j = 1,2,3$.

Finally, asymptotic normality follows by a slight adaptation of Theorem 4.1, part (a) in Hsiao (1974). ■

C.2.2 Inverse of Ω and Order of $X_1' \Omega^{-1} X_2$

C.2.2.1 Inverse of Ω

Let a matrix Ω_η have the following structure

$$\begin{aligned} \Omega_\eta &= Z^{(D)} \left(I_N \otimes J_T \otimes \tilde{\Sigma}_\lambda + J_N \otimes I_T \otimes \tilde{\Sigma}_\tau + I_{NT} \otimes \tilde{\Sigma}_\xi \right) Z^{(D)\prime} \\ &= Z^{(B)} \left(I_N \otimes \tilde{\Sigma}_\lambda \right) Z^{(B)\prime} + Z^{(T)} \left(I_T \otimes \tilde{\Sigma}_\tau \right) Z^{(T)\prime} + Z^{(D)} \left(I_{NT} \otimes \tilde{\Sigma}_\xi \right) Z^{(D)\prime} \end{aligned} \quad \text{(C.11)}$$

then we have, by the equality

$$(U + XVX')^{-1} = U^{-1} - U^{-1}X\left(V^{-1} + X'U^{-1}X\right)^{-1}X'U^{-1} \tag{C.12}$$

from Searle et al. (1992), p. 453, that

$$\Omega_\eta^{-1} = C^{-1} - C^{-1}Z^{(T)}\left\{Z^{(T)\prime}C^{-1}Z^{(T)} + \left(I_T \otimes \tilde{\Sigma}_\tau^{-1}\right)\right\}^{-1}Z^{(T)\prime}C^{-1} \tag{C.13}$$

with

$$\begin{aligned} C^{-1} &= \left[Z^{(D)}\left(I_{NT} \otimes \tilde{\Sigma}_\xi\right)Z^{(D)\prime} + Z^{(B)}\left(I_N \otimes \tilde{\Sigma}_\lambda\right)Z^{(B)\prime}\right]^{-1} \\ &= D^{-1} - D^{-1}Z^{(B)}\left\{Z^{(B)\prime}D^{-1}Z^{(B)} + \left(I_N \otimes \tilde{\Sigma}_\lambda^{-1}\right)\right\}^{-1}Z^{(B)\prime}D^{-1} \end{aligned} \tag{C.14}$$

and

$$\begin{aligned} D^{-1} &= \left[Z^{(D)}\left(I_{NT} \otimes \tilde{\Sigma}_\xi\right)Z^{(D)\prime}\right]^{-1} \\ &= \mathrm{Diag}_{i=1}^{N}\left[D_i^{-1}\right], \end{aligned} \tag{C.15}$$

where $D_j^{-1} = \mathrm{Diag}_{t=1}^{T}\left[d_{jt}^{-1}\right]$ and $d_{jt} = z_{jt}'\tilde{\Sigma}_\xi z_{jt}$.

C.2.2.2 Order of $X_1'\Omega^{-1}X_2$

Notational Issues. Let $\mathbf{D}_{LM}\left(N^lT^m;\Sigma\right)$ denote a $LM \times LM$ diagonal matrix with nonzero elements that are of order N^lT^m and that are a function of the matrix Σ. Likewise, $\mathbf{S}_{LM}\left(N^lT^m;\Sigma\right)$ denotes a $LM \times LM$ matrix, $\mathbf{B}_{LM;M}\left(N^lT^m;\Sigma\right)$ a $LM \times LM$ block-diagonal matrix with $M \times M$ blocks and $\mathbf{T}_{LM;M}\left(N^lT^m;\Sigma\right)$ a $LM \times LM$ matrix with nonzero elements where $J_L \otimes I_M$ has nonzero elements. The second argument of these matrices denotes that their nonzero elements are functions of the matrix Σ.

Lemma C.1. *The inverse (C.13) of a matrix which has a structure as given in (C.11), can be asymptotically approximated by $\Phi_A + \Phi_B + \Phi_C + \Phi_D$ in the sense that, for conformable matrices X_1 and X_2, it holds that*

$$X_1'\Omega_\eta^{-1}X_2 \xrightarrow{p} X_1'\left(\Phi_A + \Phi_B + \Phi_C + \Phi_D\right)X_2 \tag{C.16}$$

where

$$\begin{array}{rcccl}
\Phi_A & = & D^{-1} & = & \mathbf{D}_{NT}\left(1;\tilde{\Sigma}_\xi\right) \\
\Phi_B & = & -D^{-1}V^{(B)}Diag_{i=1}^{N}\left[E_i^{-1}\right]V^{(B)\prime}D^{-1} & = & \mathbf{B}_{NT;T}\left(T^{-1};\tilde{\Sigma}_\xi\right) \\
\Phi_C & = & D^{-1}V^{(T)}G_1^{-1}V^{(T)\prime}D^{-1} & = & \mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_\xi\right) \\
\Phi_D & = & D^{-1}V^{(T)}G_1^{-1}G_2G_1^{-1}V^{(T)\prime}D^{-1} & = & \mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right)
\end{array}$$

with

$$\begin{array}{rcl}
E_i & = & \sum_{s=1}^{T}\left(v_{is}'\tilde{\Sigma}_\xi v_{is}\right)^{-1}v_{is}v_{is}' \\
G_1 & = & V^{(T)\prime}D^{-1}V^{(T)} \\
G_2 & = & V^{(T)\prime}D^{-1}V^{(B)}Diag_{i=1}^{N}\left[E_i^{-1}\right]V^{(B)\prime}D^{-1}V^{(T)}.
\end{array}$$

Proof. The inverse of Ω_η^{-1} is given by (C.13). Using the above notation and (C.15), the inverse of the $KN \times KN$ matrix $V^{(B)\prime}D^{-1}V^{(B)} + I_N \otimes \tilde{\Sigma}_\lambda^{-1}$ can be written as

$$\begin{array}{rl}
 & \mathrm{Diag}_{i=1}^{N}\left[\left(\sum_{s=1}^{T}v_{is}\left(v_{is}'\tilde{\Sigma}_\xi v_{is}\right)^{-1}v_{is}' + \tilde{\Sigma}_\lambda^{-1}\right)^{-1}\right] \\
\approx & \mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right] \\
= & \mathbf{B}_{NK;K}\left(T^{-1};\tilde{\Sigma}_\xi\right)
\end{array}$$

since

$$\begin{array}{rcl}
\left(E_i + \tilde{\Sigma}_\lambda^{-1}\right)^{-1} & = & E_i^{-1} - E_i^{-1}\left(\tilde{\Sigma}_\lambda + E_i^{-1}\right)^{-1}E_i^{-1} \\
 & = & E_i^{-1} - E_i^{-1}\tilde{\Sigma}_\lambda^{-1}E_i^{-1} + E_i^{-1}\tilde{\Sigma}_\lambda^{-1}\left(\tilde{\Sigma}_\lambda^{-1} + E_i\right)^{-1}\tilde{\Sigma}_\lambda^{-1}E_i^{-1},
\end{array}$$

with $E_i = \sum_{s=1}^{T}v_{is}\left(v_{is}'\tilde{\Sigma}_\xi v_{is}\right)^{-1}v_{is}' = \mathbf{S}_K\left(T;\tilde{\Sigma}_\xi\right)$. Now it holds that $E_i^{-1} = \mathbf{S}_K\left(T^{-1};\tilde{\Sigma}_\xi\right)$, that $E_i^{-1}\tilde{\Sigma}_\lambda^{-1}E_i^{-1} = \mathbf{S}_K\left(T^{-2};\tilde{\Sigma}_\lambda,\tilde{\Sigma}_\xi\right)$ and that

$$E_i^{-1}\tilde{\Sigma}_\lambda^{-1}\left(\tilde{\Sigma}_\lambda^{-1} + E_i\right)^{-1}\tilde{\Sigma}_\lambda^{-1}E_i^{-1} \;=\; \mathbf{S}_K\left(T^{-3};\tilde{\Sigma}_\lambda,\tilde{\Sigma}_\xi\right),$$

from which we have that

$$X_1' \text{Diag}_{i=1}^N \left[\left(E_i + \tilde{\Sigma}_\lambda^{-1} \right)^{-1} \right] X_2 \quad \xrightarrow{p} \quad X_1' \text{Diag}_{i=1}^N \left[E_i^{-1} \right] X_2,$$

for which the notation $\text{Diag}_{i=1}^N \left[\left(E_i + \tilde{\Sigma}_\lambda^{-1} \right)^{-1} \right] \approx \text{Diag}_{i=1}^N \left[E_i^{-1} \right]$ is used. Consequently, it holds that

$$\begin{aligned} C^{-1} &= D^{-1} - D^{-1} V^{(B)} \left[V^{(B)\prime} D^{-1} V^{(B)} + I_N \otimes \tilde{\Sigma}_\lambda^{-1} \right]^{-1} V^{(B)\prime} D^{-1} \\ &\approx D^{-1} - D^{-1} V^{(B)} \text{Diag}_{i=1}^N \left[E_i^{-1} \right] V^{(B)\prime} D^{-1} \\ &= \mathbf{D}_{NT} \left(1; \tilde{\Sigma}_\xi \right) - \mathbf{B}_{NT;T} \left(T^{-1}; \tilde{\Sigma}_\xi \right). \end{aligned}$$

Define now $F = V^{(T)\prime} C^{-1} V^{(T)} + \left(I_T \otimes \tilde{\Sigma}_\tau^{-1} \right)$, then we have that

$$F \approx G_1 - G_2 + G_3,$$

where

$$\begin{aligned} G_1 &= V^{(T)\prime} D^{-1} V^{(T)} &&= \mathbf{B}_{TK;K} \left(N; \tilde{\Sigma}_\xi \right) \\ G_2 &= V^{(T)\prime} D^{-1} V^{(B)} \text{Diag}_{i=1}^N \left[E_i^{-1} \right] V^{(B)\prime} D^{-1} V^{(T)} &&= \mathbf{S}_{TK} \left(NT^{-1}; \tilde{\Sigma}_\xi \right) \\ G_3 &= \left(I_T \otimes \tilde{\Sigma}_\tau^{-1} \right) &&= \mathbf{B}_{TK;K} \left(1; \tilde{\Sigma}_\tau \right) \end{aligned}$$

and

$$\begin{aligned} F^{-1} &\approx (G_1 - G_2 + G_3)^{-1} \\ &= G_1^{-1} + G_1^{-1} G_2 G_1^{-1} - G_1^{-1} G_3 G_1^{-1} \\ &\approx G_1^{-1} + G_1^{-1} G_2 G_1^{-1} \\ &= \mathbf{B}_{TK;K} \left(N^{-1}; \tilde{\Sigma}_\xi \right) + \mathbf{S}_{TK} \left(N^{-1} T^{-1}; \tilde{\Sigma}_\xi \right), \end{aligned}$$

by applying (C.12) and because $G_1^{-1} G_3 G_1^{-1} = \mathbf{B}_{TK;K} \left(N^{-2}; \tilde{\Sigma}_\tau, \tilde{\Sigma}_\xi \right)$. Consequently, it

holds that

$$\begin{aligned} C^{-1}V^{(T)}F^{-1}V^{(T)\prime}C^{-1} &\approx C^{-1}V^{(T)}G_1^{-1}V^{(T)\prime}C^{-1} \\ &\quad +C^{-1}V^{(T)}G_1^{-1}G_2G_1^{-1}V^{(T)\prime}C^{-1} \\ &= \mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_\xi\right)+\mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right) \end{aligned}$$

and the inverse covariance matrix can be written as

$$\begin{aligned} \Omega_\eta^{-1} &= C^{-1}-C^{-1}V^{(T)}F^{-1}V^{(T)\prime}C^{-1} \\ &\approx D^{-1}-D^{-1}V^{(B)}\mathrm{Diag}_{i=1}^{N}\left[E_i^{-1}\right]V^{(B)\prime}D^{-1} \\ &\quad +D^{-1}V^{(T)}G_1^{-1}V^{(T)\prime}D^{-1} \\ &\quad +D^{-1}V^{(T)}G_1^{-1}G_2G_1^{-1}V^{(T)\prime}D^{-1} \\ &= \mathbf{D}_{NT}\left(1;\tilde{\Sigma}_\xi\right)-\mathbf{B}_{NT;T}\left(T^{-1};\tilde{\Sigma}_\xi\right) \\ &\quad +\mathbf{T}_{NT;T}\left(N^{-1};\tilde{\Sigma}_\xi\right)+\mathbf{S}_{NT}\left(N^{-1}T^{-1};\tilde{\Sigma}_\xi\right), \end{aligned}$$

which proves the lemma. ■

Remark. Note that $X_1\Phi_J X_2 = O(NT)$ for $J = A, B, C, D$, for all conformable matrices, for which $X_1 = O_p(1)$ and $X_2 = O_p(1)$. Note also that the leading terms in the asymptotic approximation of Ω_η^{-1} are all functions of $\tilde{\Sigma}_\xi$, but neither of $\tilde{\Sigma}_\lambda$ nor $\tilde{\Sigma}_\tau$.

C.2.3 Aitken Estimation

Proof of Theorem 4.2. Consistency of

$$\hat{\varphi}_{IVA} = \left(R'\mathcal{P}_V\Omega_\eta^{-1}R\right)^{-1}R'\mathcal{P}_V\Omega_\eta^{-1}Y$$

follows from the consistency of $\hat{\varphi}_{IV}$. Following Lemma C.1 (see the remark above), we have that $P_\Omega = \mathrm{plim}_{N,T\to\infty}\left[(NT)^{-1}\hat{R}'\Omega_\eta^{-1}R\right] = 0$, since both matrices $\hat{R}$ and R are $O_p(1)$, it holds that $P_\Omega = O_p(1)$ Furthermore, for $\Gamma_\Omega = \lim_{N,T\to\infty}\left[(NT)^{-2}\Psi_\Omega\hat{R}'\Omega_\eta^{-1}\hat{R}\Psi_\Omega\right]$ to be $O(1)$, we need that $N_j = NT$ for $j = i, p, r$. This proves the consistency rate in the general case. Asymptotic normality again follows by a slight adaptation of Theorem 4.1. in Hsiao (1974).

To prove the case of an error component model, we notice that Ω_η can be written as

(Wyhowski (1998))

$$\Omega_\eta = \left(T\sigma_\lambda^2 + \sigma_\xi^2\right)\mathcal{P}_i + \left(N\sigma_\tau^2 + \sigma_\xi^2\right)\mathcal{P}_p + \sigma_\xi^2\mathcal{P}_r,$$

and its inverse as

$$\Omega_\eta^{-1} = \left(T\sigma_\lambda^2 + \sigma_\xi^2\right)^{-1}\mathcal{P}_i + \left(N\sigma_\tau^2 + \sigma_\xi^2\right)^{-1}\mathcal{P}_p + \sigma_\xi^{-2}\mathcal{P}_r.$$

Therefore, it holds that

$$\begin{aligned}(NT)^{-1} R'\mathcal{P}_V\Omega_\eta^{-1}R &= \left(T\sigma_\lambda^2 + \sigma_\xi^2\right)^{-1}(NT)^{-1} R'\mathcal{P}_{V_i}R \\ &\quad + \left(N\sigma_\tau^2 + \sigma_\xi^2\right)^{-1}(NT)^{-1} R'\mathcal{P}_{V_p}R \\ &\quad + \sigma_\xi^{-2}(NT)^{-1} R'\mathcal{P}_{V_r}R \end{aligned} \tag{C.17}$$

and

$$\begin{aligned}(NT)^{-2}\Psi_\Omega R'\mathcal{P}_V\Omega_\eta^{-1}\mathcal{P}_V R\Psi_\Omega &= \left(T\sigma_\lambda^2 + \sigma_\xi^2\right)^{-1} N_i (NT)^{-2} R'\mathcal{P}_{V_i}R \\ &\quad + \left(N\sigma_\tau^2 + \sigma_\xi^2\right)^{-1} N_p (NT)^{-2} R'\mathcal{P}_{V_p}R \\ &\quad + \sigma_\xi^{-2} N_r (NT)^{-2} R'\mathcal{P}_{V_r}R. \end{aligned} \tag{C.18}$$

For individual-specific variables we have that

$$\begin{aligned}&\operatorname*{plim}_{N,T\to\infty}\left[(NT)^{-1} R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_V\Omega_\eta^{-1}R^{\ddot{b}\ddot{c}}\right] \\ &= \lim_{N,T\to\infty}\left[T^{-1}\sigma_\lambda^{-2}\right]\operatorname*{plim}_{N,T\to\infty}\left[(NT)^{-1} R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{V_i}R^{\ddot{b}\ddot{c}}\right] \\ &= O\left(T^{-1}\right),\end{aligned}$$

$$\begin{aligned}&\lim_{N,T\to\infty}\left[(NT)^{-2}\Psi_\Omega R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_V\Omega_\eta^{-1}\mathcal{P}_V R^{\ddot{b}\ddot{c}}\Psi_\Omega\right] \\ &= \lim_{N,T\to\infty}\left[N_i N^{-1}T^{-2}\sigma_\lambda^{-2}\right]\operatorname*{plim}_{N,T\to\infty}\left[(NT)^{-1} R^{\ddot{b}\ddot{c}\prime}\mathcal{P}_{V_i}R^{\ddot{b}\ddot{c}}\right] \\ &= N_i O\left(N^{-1}T^{-2}\right)\end{aligned}$$

and $N_i = N$. Similarly, $N_p = T$ and $N_r = NT$. ■

It can also be seen from (C.17) and (C.18) that variables containing more than one

component have as consistency rate the highest (fastest) consistency rate of each of the components. For example, consider a variable w_{it} with both an individual-specific and a period-specific component and assume that $N = o(T)$, which implies $T^{-1} = o(N^{-1})$. Now $(NT)^{-1} W'\mathcal{P}_V \Omega_\eta^{-1} W$ will be the sum of two terms: $O(T^{-1}) + O(N^{-1})$, of which asymptotically $O(N^{-1})$ (the period-specific component) will dominate. The period specific component of $(NT)^{-2} W'\mathcal{P}_V \Omega_\eta^{-1} \mathcal{P}_V W$, will also dominate. Variables with both individual-specific and period-specific components will thus have the same rate of convergence as period-specific variables: $\sqrt{T}$.

C.2.4 Covariance Estimation

Proof of Proposition 4.2. By construction of $V^{(T)}$ and $E^{(T)}$, we have that

$$\operatorname*{plim}_{N,T\to\infty} \left[V^{(T)\prime} V^{(T)}\right] = \begin{pmatrix} O(NT^2) & O(NT) & O(NT) \\ O(NT) & O(N^2T) & O(NT) \\ O(NT) & O(NT) & O(NT) \end{pmatrix}$$

and

$$\operatorname*{plim}_{N,T\to\infty} \left[\left(V^{(T)\prime} V^{(T)}\right)^{-1}\right] = \begin{pmatrix} O(N^{-1}T^{-2}) & O(N^{-2}T^{-2}) & O(N^{-1}T^{-2}) \\ O(N^{-2}T^{-2}) & O(N^{-2}T^{-1}) & O(N^{-2}T^{-1}) \\ O(N^{-1}T^{-2}) & O(N^{-2}T^{-1}) & O(N^{-1}T^{-1}) \end{pmatrix}. \quad \text{(C.19)}$$

The matrix $\operatorname{Var}\left[V^{(T)\prime} E^{(T)}\right]$ will now be computed.

From the definition $H_j^{(\lambda)} = (\hat{\eta}_{j1}\hat{\eta}_{j2}, \hat{\eta}_{j1}\hat{\eta}_{j3}, \ldots, \hat{\eta}_{j,T-1}\hat{\eta}_{jT})'$ as a $T(T-1)/2 \times 1$-matrix, we have that $\mathrm{E}\left[H_i^{(\lambda)} H_j^{(\lambda)\prime}\right] - \mathrm{E}\left[H_i^{(\lambda)}\right] \mathrm{E}\left[H_j^{(\lambda)\prime}\right]$ has a st, uv-th element equal to

$$\begin{aligned} c_{ij;st,uv} = {} & \left(v_{is}' \otimes v_{jv}'\right) \mathrm{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_t' + \xi_{is}\xi_{it}'\right) \otimes \left(\lambda_j\lambda_j' + \tau_v\tau_u' + \xi_{jv}\xi_{ju}'\right)\right] \left(v_{it} \otimes v_{ju}\right) \\ & - v_{is}' \left\{\Sigma_\lambda + \delta_{s=t}\left(\Sigma_\tau + \Sigma_\xi\right)\right\} v_{it} \cdot v_{jv}' \left\{\Sigma_\lambda + \delta_{u=v}\left(\Sigma_\tau + \Sigma_\xi\right)\right\} v_{ju}. \end{aligned}$$

Now it holds that

$$\begin{aligned} & \left(v_{is}' \otimes v_{jv}'\right) \mathrm{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_t' + \xi_{is}\xi_{it}'\right) \otimes \left(\lambda_j\lambda_j' + \tau_v\tau_u' + \xi_{jv}\xi_{ju}'\right)\right] \left(v_{it} \otimes v_{ju}\right) \\ = {} & v_{is}' \left\{\Sigma_\lambda + \delta_{s=t}\left(\Sigma_\tau + \Sigma_\xi\right)\right\} v_{it} \cdot v_{jv}' \left\{\Sigma_\lambda + \delta_{u=v}\left(\Sigma_\tau + \Sigma_\xi\right)\right\} v_{ju} \end{aligned}$$

if $i \neq j$, $s \neq u$ and $t \neq v$. From this we deduce that

$$\mathrm{E}\left[H_i^{(\lambda)} H_j^{(\lambda)\prime}\right] - \mathrm{E}\left[H_i^{(\lambda)}\right] \mathrm{E}\left[H_j^{(\lambda)\prime}\right] \;=\; \mathrm{Diag}_{s=1}^{(T-1)}\left[\mathrm{Diag}_{t=s+1}^{T}\left[c_{ij;st,st}\right]\right],$$

if $i \neq j$, and it is equal to a $T(T-1)/2 \times T(T-1)/2$ matrix with nonzero elements if $i = j$. Thus it holds that

$$\begin{aligned}
\mathrm{Var}\left[V^{(\lambda)\prime} E^{(\lambda)}\right] &= \mathrm{Var}\left[V^{(\lambda)\prime}\left(H^{(\lambda)} - \mathrm{E}\left[H^{(\lambda)}\right]\right)\right] \\
&= V^{(\lambda)\prime}\left(\mathrm{E}\left[H^{(\lambda)} H^{(\lambda)\prime}\right] - \mathrm{E}\left[H^{(\lambda)}\right] \mathrm{E}\left[H^{(\lambda)\prime}\right]\right) V^{(\lambda)} \\
&= O\left(N(N-1)T(T-1)/2\right) + O\left(NT^2(T-1)^2/4\right).
\end{aligned}$$

This is the sum of N blocks for which $i = j$, which contain $T^2(T-1)^2/4$ nonzero elements each, plus the $N(N-1)$ blocks for which $i \neq j$, which contain $T(T-1)/2$ nonzero elements. Since the problem at hand is symmetric with respect to N and T, we also have that

$$\mathrm{Var}\left[V^{(\tau)\prime} E^{(\tau)}\right] = O\left(N(N-1)T(T-1)/2\right) + O\left(N^2(N-1)^2 T/4\right).$$

Similarly, the s,t-th element of $\mathrm{E}\left[H_i^{(\xi)} H_j^{(\xi)\prime}\right] - \mathrm{E}\left[H_i^{(\xi)}\right] \mathrm{E}\left[H_j^{(\xi)\prime}\right]$ equals

$$\begin{aligned}
d_{ij;st} \;=\; & \left(v_{is}' \otimes v_{jt}'\right) \mathrm{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_s' + \xi_{is}\xi_{is}'\right) \otimes \left(\lambda_j\lambda_j' + \tau_t\tau_t' + \xi_{jt}\xi_{jt}'\right)\right] \left(v_{is} \otimes v_{jt}\right) \\
& - v_{is}'\left\{\Sigma_\lambda + \Sigma_\tau + \Sigma_\xi\right\} v_{is} \cdot v_{jt}'\left\{\Sigma_\lambda + \Sigma_\tau + \Sigma_\xi\right\} v_{jt},
\end{aligned}$$

which is equal to zero if $i \neq j$ and $s \neq t$. As a result it holds that

$$\begin{aligned}
\mathrm{Var}\left[V^{(\xi)\prime} E^{(\xi)}\right] &= V^{(\xi)\prime}\left(\mathrm{E}\left[H^{(\xi)} H^{(\xi)\prime}\right] - \mathrm{E}\left[H^{(\xi)}\right] \mathrm{E}\left[H^{(\xi)\prime}\right]\right) V^{(\xi)} \\
&= O\left(NT^2 + N(N-1)T\right),
\end{aligned}$$

which is the sum of N blocks for which $i = j$, containing T^2 elements, and $N(N-1)$ blocks for wich $i \neq j$, containing T elements.
Furthermore, $\mathrm{E}\left[H_i^{(\lambda)} H_j^{(\xi)\prime}\right] - \mathrm{E}\left[H_i^{(\lambda)}\right] \mathrm{E}\left[H_j^{(\xi)\prime}\right]$, with st,u-th element

$$\begin{aligned}
e_{ij;st,u} \;=\; & \left(v_{is}' \otimes v_{ju}'\right) \mathrm{E}\left[\left(\lambda_i\lambda_i' + \tau_s\tau_t' + \xi_{is}\xi_{it}'\right) \otimes \left(\lambda_j\lambda_j' + \tau_u\tau_u' + \xi_{ju}\xi_{ju}'\right)\right] \left(v_{it} \otimes v_{ju}\right) \\
& - v_{is}'\left\{\Sigma_\lambda\right\} v_{it} \cdot v_{ju}'\left\{\Sigma_\lambda + \Sigma_\tau + \Sigma_\xi\right\} v_{ju},
\end{aligned}$$

is equal to $\mathbf{O}_{T(T-1)\times T}$, unless $i = j$, in which case all of its elements are nonzero. Consequently we have that $\mathrm{Cov}\left[V^{(\lambda)\prime} E^{(\lambda)}, V^{(\xi)\prime} E^{(\xi)}\right] = O\left(NT^2(T-1)\right)$, and, by the symmetry

between N and T, $\text{Cov}\left[V^{(\tau)\prime}E^{(\tau)}, V^{(\xi)\prime}E^{(\xi)}\right] = O\left(N^2\left(N-1\right)T\right)$. Finally, $\text{E}\left[H_i^{(\lambda)}H_t^{(\tau)\prime}\right] - \text{E}\left[H_i^{(\lambda)}\right]\text{E}\left[H_t^{(\tau)\prime}\right]$ has uv,kl-th element equal to

$$\begin{aligned} f_{it;uv,kl} &= \left(v_{iu}'\otimes v_{kt}'\right)\text{E}\left[\left(\lambda_i\lambda_i' + \tau_u\tau_v' + \xi_{iu}\xi_{iv}'\right)\otimes\left(\lambda_k\lambda_l' + \tau_t\tau_t' + \xi_{kt}\xi_{lt}'\right)\right]\left(v_{iv}\otimes v_{lt}\right) \\ &\quad -v_{iu}'\left\{\Sigma_\lambda\right\}v_{iv}\cdot v_{kt}'\left\{\Sigma_\tau\right\}v_{lt}, \end{aligned}$$

which is equal to zero, unless $\left(h=i\vee l=i\right)\wedge\left(u=t\vee v=t\right)$, i.e. at $\left(2N-3\right)\left(2T-3\right)$ places. It follows that $\text{Cov}\left[Z^{(\lambda)\prime}E^{(\lambda)}, Z^{(\tau)\prime}E^{(\tau)}\right] = O\left(N\left(2N-3\right)T\left(2T-3\right)\right)$.

From the previous we have that

$$\begin{aligned} &\text{Var}\left[V^{(T)\prime}E^{(T)}\right] \\ &= \begin{pmatrix} O\left(NT^4\right)+O\left(N^2T^2\right) & O\left(N^3T\right)+O\left(NT^3\right) & O\left(NT^3\right)+O\left(N^2T\right) \\ O\left(N^3T\right)+O\left(NT^3\right) & O\left(N^4T\right)+O\left(N^2T^2\right) & O\left(N^3T\right)+O\left(NT^2\right) \\ O\left(NT^3\right)+O\left(N^2T\right) & O\left(N^3T\right)+O\left(NT^2\right) & O\left(N^2T\right)+O\left(NT^2\right) \end{pmatrix}, \end{aligned}$$

and the diagonal elements of

$$\operatorname*{plim}_{N,T\to\infty}\left[\left(V^{(T)\prime}V^{(T)}\right)^{-1}\right]\text{Var}\left[V^{(T)\prime}E^{(T)}\right]\operatorname*{plim}_{N,T\to\infty}\left[\left(V^{(T)\prime}V^{(T)}\right)^{-1}\right]$$

are given by $\left(O\left(N^{-1}\right)+O\left(T^{-2}\right), O\left(N^{-2}\right)+O\left(T\right), O\left(N^{-1}\right)+O\left(T^{-1}\right)\right)$, which proves the proposition. ■

C.2.5 Feasible Aitken Estimation

The proof of Theorem 4.3 consists of two parts. The first part proves the equivalence of Aitken and feasible Aitken estimators in the case we are dealing with an error-components model. The second part proves the rate of convergence of the variance in the general case.

Proof of Theorem 4.3. To prove the first part of the Theorem, we need to prove that

$$\begin{aligned} \operatorname*{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1}\hat{R}'\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)R\right] &= 0 \\ \operatorname*{plim}_{N,T\to\infty}\left[\left(NT\right)^{-1}\Psi_\Omega\hat{R}'\left(\hat{\Omega}_\eta^{-1}-\Omega_\eta^{-1}\right)H\right] &= 0. \end{aligned}$$

We have, by the consistency of $\left(\hat{\sigma}_\lambda^2, \hat{\sigma}_\tau^2, \hat{\sigma}_\xi^2\right)$, that

$$\begin{aligned}
\underset{N,T\to\infty}{\operatorname{plim}} \left[(NT)^{-1} \hat{R}' \left(\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\right) R\right] &= \underset{N,T\to\infty}{\operatorname{plim}} \left[(NT)^{-1} \hat{R}' \mathcal{P}_i R \left(\frac{\sigma_\lambda^2 - \hat{\sigma}_\lambda^2}{\hat{\sigma}_\lambda^2 \sigma_\lambda^2}\right)\right] \\
&\quad + \underset{N,T\to\infty}{\operatorname{plim}} \left[(NT)^{-1} \hat{R}' \mathcal{P}_p R \left(\frac{\sigma_\tau^2 - \hat{\sigma}_\tau^2}{\hat{\sigma}_\tau^2 \sigma_\tau^2}\right)\right] \\
&\quad + \underset{N,T\to\infty}{\operatorname{plim}} \left[(NT)^{-1} \hat{R}' \mathcal{P}_r R \left(\frac{\sigma_\xi^2 - \hat{\sigma}_\xi^2}{\hat{\sigma}_\xi^2 \sigma_\xi^2}\right)\right] \\
&= o(1),
\end{aligned}$$

and that

$$\begin{aligned}
(NT)^{-1} \Psi_\Omega \hat{R}' \hat{\Omega}_\eta^{-1} H &= (NT)^{-1} \left(T\hat{\sigma}_\lambda^2\right)^{-1} \Psi_\Omega \hat{R}' \mathcal{P}_i H \\
&\quad + (NT)^{-1} \left(N\hat{\sigma}_\tau^2\right)^{-1} \Psi_\Omega \hat{R}' \mathcal{P}_p H \\
&\quad + (NT)^{-1} \hat{\sigma}_\xi^{-2} \Psi_\Omega \hat{R}' \mathcal{P}_r H \\
&= T_1 + T_2 + T_3 + T_4,
\end{aligned}$$

where

$$\begin{aligned}
T_1 &= (NT)^{-1} \Psi_\Omega \hat{R}' \Omega_\eta^{-1} H \\
T_2 &= \sqrt{N_i} N^{-1} T^{-2} \sigma_\lambda^{-4} \left(\sigma_\lambda^2 - \hat{\sigma}_\lambda^2\right) \hat{R}' \mathcal{P}_i H \\
T_3 &= \sqrt{N_p} N^{-2} T^{-1} \sigma_\tau^{-4} \left(\sigma_\tau^2 - \hat{\sigma}_\tau^2\right) \hat{R}' \mathcal{P}_p H \\
T_4 &= \sqrt{N_r} (NT)^{-1} \sigma_\xi^{-4} \left(\sigma_\xi^2 - \hat{\sigma}_\xi^2\right) \hat{R}' \mathcal{P}_r H
\end{aligned}$$

The first term, T_1, is the numerator of the Aitken estimator (see Theorem 4.2) and it has variance equal to $O(1)$. The second term can be written as

$$\begin{aligned}
T_2 &= \sqrt{N_i} N^{-1} T^{-2} \sigma_\lambda^{-4} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \hat{r}_{it} \left(\lambda_i + \frac{\xi_{it}}{T}\right) \right. \\
&\quad \left. \times \frac{2}{NT(T-1)} \sum_{j=1}^{N} \sum_{r=1}^{T} \sum_{s=r+1}^{T} \left(\eta_{jr}\eta_{js} - \mathrm{E}\left[\eta_{jr}\eta_{js}\right]\right) \right\},
\end{aligned}$$

where the second line is equal to $\sigma_\lambda^2 - \hat{\sigma}_\lambda^2$, with $\hat{\sigma}_\lambda^2$ the covariance estimator from subsection

4.4.3 in the case of an error component model. T_2 has variance equal to

$$\begin{aligned}
\mathrm{Var}\left[T_2\right] &= \frac{4\sigma_\lambda^{-8}N_i}{N^4T^6\left(T-1\right)^2}\sum_{i,j,k,l}^{N}\sum_{s,t,u,w}^{T}\sum_{v>u,x>w}^{T}\mathrm{E}\left[\hat{r}_{it}\left(\lambda_i+\frac{\xi_{it}}{T}\right)\left(\eta_{ku}\eta_{kv}-\mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\right.\\
&\qquad\left.\times\left(\eta_{lw}\eta_{lx}-\mathrm{E}\left[\eta_{lw}\eta_{lx}\right]\right)\left(\lambda_j+\frac{\xi_{js}}{T}\right)\hat{r}'_{js}\right]\\
&\quad-\frac{4\sigma_\lambda^{-8}N_i}{N^4T^6\left(T-1\right)^2}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\sum_{v>u}^{T}\mathrm{E}\left[\hat{r}_{it}\left(\lambda_i+\frac{\xi_{it}}{T}\right)\left(\eta_{ku}\eta_{kv}-\mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\right]\right)^2\\
&= N_iO\left(N^{-2}T^{-2}\right)J_{(1+K)\left(1+\frac{K}{2}+J\right)},
\end{aligned}$$

since

$$\begin{aligned}
&\sum_{i,j,k,l}^{N}\sum_{s,t,u,v,w,x}^{T}\mathrm{E}\left[\left(\eta_{ku}\eta_{kv}-\mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\left(\lambda_i+\frac{\xi_{it}}{T}\right)\left(\lambda_j+\frac{\xi_{js}}{T}\right)\left(\eta_{lw}\eta_{lx}-\mathrm{E}\left[\eta_{lw}\eta_{lx}\right]\right)\right]\\
-&\left(\sum_{i,k}^{N}\sum_{t,u,v}^{T}\mathrm{E}\left[\left(\eta_{ku}\eta_{kv}-\mathrm{E}\left[\eta_{ku}\eta_{kv}\right]\right)\left(\lambda_i+\frac{\xi_{it}}{T}\right)\right]\right)^2\\
=\;& O\left(N^2T^6\right),
\end{aligned}$$

because the LHS is equal to zero if $i\neq j$ or $k\neq l$. Similarly we have that $\mathrm{Var}\left[T_3\right]=N_pO\left(N^{-2}T^{-2}\right)J_{(1+K)\left(1+\frac{K}{2}+J\right)}$ and that

$$\begin{aligned}
\mathrm{Var}\left[T_4\right] &= \frac{\sigma_\xi^{-8}}{N^4T^4}\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\hat{r}_{it}\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\xi_{it}\right.\\
&\qquad\left.\times\xi_{js}\left(\eta_{lv}^2-\mathrm{E}\left[\eta_{lv}^2\right]\right)\hat{r}'_{js}\right]\\
&\quad-\frac{\sigma_\xi^{-8}}{N^4T^4}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\hat{r}_{it}\left(\eta_{ku}^2-\mathrm{E}\left[\eta_{ku}^2\right]\right)\xi_{it}\right]\right)^2\\
&= N_rO\left(N^{-2}T^{-2}\right)J_{(1+K)\left(1+\frac{K}{2}+J\right)}.
\end{aligned}$$

Since we have, from Theorem 4.2, that $\left(\begin{array}{cccc} N_i, & N_p, & N_r \end{array}\right)=\left(\begin{array}{ccc} N, & T, & NT \end{array}\right)$, the above expressions prove that

$$\begin{aligned}
\left(NT\right)^{-1}\hat{R}'\hat{\Omega}_\eta^{-1}H &\xrightarrow{p} 0,\\
\left(NT\right)^{-1}\Psi_\Omega\hat{R}'\hat{\Omega}_\eta^{-1}H &\xrightarrow{d} \left(NT\right)^{-1}\Psi_\Omega\hat{R}'\Omega_\eta^{-1}H,
\end{aligned}$$

and thus the first part of the theorem.

The second part of the theorem states that $\min(N,T)(\hat{\varphi}_{FIVA}-\varphi)=O_p(1)$ as $N,T\to\infty$. To prove it, note that

$$\begin{aligned}(NT)^{-1}\hat{R}'\hat{\Omega}_\eta^{-1}W \xrightarrow{p} & (NT)^{-1}\hat{R}'\hat{D}^{-1}W \\ & -(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(B)}\mathrm{Diag}_{i=1}^{N}\left[\hat{E}_i^{-1}\right]V^{(B)\prime}\hat{D}^{-1}W \\ & +(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(T)}\hat{G}_1^{-1}V^{(T)\prime}\hat{D}^{-1}W \\ & +(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(T)}\hat{G}_1^{-1}\hat{H}_1\hat{G}_1^{-1}V^{(T)\prime}\hat{D}^{-1}W, \end{aligned} \tag{C.20}$$

with $W=R,H$, by Lemma C.1.

First it holds that $(NT)^{-1}\hat{R}'\hat{\Omega}_\eta^{-1}R=O_p(1)$, because $(NT)^{-1}\hat{R}'\hat{D}^{-1}R$ has as p,q-th element

$$\begin{aligned}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;p}\left(v_{it}'\hat{\tilde{\Sigma}}_\xi v_{it}\right)^{-1}r_{it;q} &= (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;p}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-1}r_{it;q} \\ &\quad +(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\Big\{\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2}v_{it}^{(\xi)\prime} \\ &\qquad \times\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1}\left[Z^{(\xi)\prime}E^{(\xi)}\right] \\ &\qquad +O_p(\mathbf{R}_r)\Big\}r_{it;q},\end{aligned}$$

with $O_p(\mathbf{R}_r)=O_p\left(\sum_{j,k}^{k_\eta}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_j\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_k\right)$, since

$$\begin{aligned}\left(v_{it}'\hat{\tilde{\Sigma}}_\xi v_{it}\right)^{-1} &= \left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-1}+\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2}v_{it}^{(\xi)\prime}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right) \\ &\quad +O_p\left(\sum_{j,k}^{k_\eta}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_j\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_k\right),\end{aligned}$$

by (4.19), using the definitions of $\tilde{\Sigma}^{(\xi)}$ and $v_{it}^{(\xi)}$ from subsection 4.4.3. We further have that

$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;p}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-1}r_{it;q}=O_p\left(1\right)$, and that

$$(NT)^{-1}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}\left\{\hat{r}_{it;p}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-2}v_{it}^{(\xi)\prime}\times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}v_{js}^{(\xi)}\left(\eta_{js}^{2}-\mathrm{E}\left[\eta_{js}^{2}\right]\right)r_{it;q}\right\} \;=\; o_p\left(1\right),$$

which proves $(NT)^{-1}\hat{R}'\hat{D}^{-1}R=O_p\left(1\right)$. The other terms in (C.20) are of of the same order in probability as $(NT)^{-1}\hat{R}'\hat{D}^{-1}R$ and hence $(NT)^{-1}\hat{R}'\hat{\Omega}_{\eta}^{-1}R=O_p\left(1\right)$.

Secondly, we have that $(NT)^{-1}\hat{R}'\hat{D}^{-1}H$ has as q^{th} element

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left(v_{it}'\hat{\tilde{\Sigma}}_{\xi}v_{it}\right)^{-1}\eta_{it} \;=\; T_1+T_2+T_3+T_4, \tag{C.21}$$

where

$$T_1 \;=\; (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-1}\eta_{it},$$

$$T_2 \;=\; (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left\{\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-2}v_{it}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}\left[V^{(\xi)\prime}E^{(\xi)}\right]\right\}\eta_{it},$$

$$\begin{aligned}T_3 \;=\; (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\Bigg\{&\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-3}v_{it}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}\left[V^{(\xi)\prime}E^{(\xi)}\right]\\&\times\left[E^{(\xi)\prime}V^{(\xi)}\right]\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}v_{it}^{(\xi)}\Bigg\}\eta_{it},\end{aligned}$$

$$T_4 \;=\; (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left\{O_p\left(\mathbf{R}_{\eta}\right)\right\}\eta_{it}$$

and

$$O_p\left(\mathbf{R}_{\eta}\right) \;=\; O_p\left(\sum_{j,k,l}^{k_{\eta}}\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_j\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_k\left(\hat{\tilde{\Sigma}}^{(\xi)}-\tilde{\Sigma}^{(\xi)}\right)_l\right),$$

since

$$\begin{aligned}\left(v_{it}'\hat{\tilde{\Sigma}}_\xi v_{it}\right)^{-1} &= \left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-1} + \left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} v_{it}^{(\xi)\prime}\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right) \\ &\quad + \left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-3} v_{it}^{(\xi)\prime}\left(\left[\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right]\left[\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right]'\right) v_{it}^{(\xi)} \\ &\quad + O_p\left(\sum_{j,k,l}^{k_\eta}\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right)_j\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right)_k\left(\hat{\tilde{\Sigma}}^{(\xi)} - \tilde{\Sigma}^{(\xi)}\right)_l\right),\end{aligned}$$

by (4.19), using the definitions of $\tilde{\Sigma}^{(\xi)}$ and $v_{it}^{(\xi)}$ from subsection 4.4.3. The first term of (C.21), $T_1 = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-1}\eta_{it}$, has variance equal to $O_p\left((NT)^{-1}\right)$. The second term can be written as

$$\begin{aligned}T_2 &= (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} v_{it}^{(\xi)\prime}\right. \\ &\qquad \left.\times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T} v_{js}^{(\xi)}\left(\eta_{js}^2 - \mathrm{E}\left[\eta_{js}^2\right]\right)\eta_{it}\right\},\end{aligned}$$

and its variance has q^{th} diagonal element equal to

$$\begin{aligned}\mathrm{Var}\left[T_2\right] &= (NT)^{-2}\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} v_{it}^{(\xi)\prime}\right. \\ &\qquad \times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it} \\ &\qquad \times\eta_{js}\left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right) v_{lv}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} \\ &\qquad \left.\times v_{ku}^{(\xi)}\left(v_{js}'\tilde{\Sigma}_\xi v_{js}\right)^{-2}\hat{r}_{js;q}\right] \\ &\quad - (NT)^{-2}\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} v_{it}^{(\xi)\prime}\right.\right. \\ &\qquad \left.\left.\times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right)^2 \\ &= O\left(N^{-2} + T^{-2}\right),\end{aligned}$$

since

$$\begin{aligned}
&\sum_{i,j,k,l}^{N}\sum_{s,t,u,v}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\eta_{js}\left(\eta_{lv}^{2}-\mathrm{E}\left[\eta_{lv}^{2}\right]\right)\right]\\
&-\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\right]\right)^{2}\\
=\;& O\left(N^{2}T^{4}+N^{4}T^{2}\right), \qquad \text{(C.22)}
\end{aligned}$$

and since $V^{(\xi)\prime}V^{(\xi)}=O\left(NT\right)$. Similarly, the q,r-th element of $\mathrm{Cov}\left[T_1T_2\right]$ is given by

$$\begin{aligned}
\mathrm{Cov}\left[T_2T_1\right] =\;& \left(NT\right)^{-2}\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-2}v_{it}^{(\xi)\prime}\right.\\
&\times\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1}v_{ku}^{(\xi)}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\\
&\left.\times\eta_{js}\left(v_{js}'\tilde{\Sigma}_{\xi}v_{js}\right)^{-1}\hat{r}_{js;r}\right]\\
=\;& O\left(N^{-2}+T^{-2}\right),
\end{aligned}$$

since

$$\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\eta_{js}\right] = O\left(NT^{3}+N^{3}T\right).$$

T_3 can be written as

$$\begin{aligned}
T_3 =\;& \left(NT\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-3}v_{it}^{(\xi)\prime}\right.\\
&\times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}v_{js}^{(\xi)}\left(\eta_{js}^{2}-\mathrm{E}\left[\eta_{js}^{2}\right]\right)\\
&\left.\times\sum_{k=1}^{N}\sum_{u=1}^{T}\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)v_{ku}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1}v_{it}^{(\xi)}\eta_{it}\right\},
\end{aligned}$$

with q^{th} diagonal element of its variance

$$\begin{aligned}
\text{Var}\left[T_3\right] &= (NT)^{-2} \sum_{i,j,k,l,m,n}^{N} \sum_{s,t,u,v,x,y}^{T} \text{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-3} v_{it}^{(\xi)\prime}\right. \\
&\quad \times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \text{E}\left[\eta_{ku}^2\right]\right) \\
&\quad \times\left(\eta_{lv}^2 - \text{E}\left[\eta_{lv}^2\right]\right) v_{lv}^{(\xi)\prime}\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} \\
&\quad \times v_{it}^{(\xi)}\eta_{it}\eta_{js}v_{js}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} \\
&\quad \times z_{mx}^{(\xi)}\left(\eta_{mx}^2 - \text{E}\left[\eta_{mx}^2\right]\right)\left(\eta_{ny}^2 - \text{E}\left[\eta_{ny}^2\right]\right) v_{ly}^{(\xi)\prime} \\
&\quad \left.\times\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{js}^{(\xi)}\left(v_{js}'\tilde{\Sigma}_{\xi}v_{js}\right)^{-3}\hat{r}_{js;q}\right] \\
&\quad - (NT)^{-2}\left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\text{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_{\xi}v_{it}\right)^{-3} v_{it}^{(\xi)\prime}\right.\right. \\
&\quad \times\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{js}^{(\xi)}\left(\eta_{js}^2 - \text{E}\left[\eta_{js}^2\right]\right) \\
&\quad \left.\left.\times\left(\eta_{ku}^2 - \text{E}\left[\eta_{ku}^2\right]\right) v_{ku}^{(\xi)\prime}\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{it}^{(\xi)}\eta_{it}\right]\right)^2 \\
&= O\left(N^{-3} + T^{-3}\right),
\end{aligned}$$

since

$$\begin{aligned}
&\sum_{i,j,k,l,m,n}^{N}\sum_{s,t,u,v,x,y}^{T}\text{E}\left[\left(\eta_{ku}^2 - \text{E}\left[\eta_{ku}^2\right]\right)\left(\eta_{lv}^2 - \text{E}\left[\eta_{lv}^2\right]\right)\right. \\
&\qquad \left.\times\eta_{it}\eta_{js}\left(\eta_{mx}^2 - \text{E}\left[\eta_{mx}^2\right]\right)\left(\eta_{ny}^2 - \text{E}\left[\eta_{ny}^2\right]\right)\right] \\
&\quad - \left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\text{E}\left[\left(\eta_{js}^2 - \text{E}\left[\eta_{js}^2\right]\right)\left(\eta_{ku}^2 - \text{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right)^2 \\
&= O\left(N^3T^6 + N^6T^3\right).
\end{aligned}$$

The q,r-th element of $\mathrm{Cov}\left[T_3 T_1\right]$ is given by

$$\begin{aligned}
\mathrm{Cov}\left[T_3, T_1\right] &= (NT)^{-2} \sum_{i,j,k,l}^{N} \sum_{s,t,u,v}^{T} \mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-3} v_{it}^{(\xi)\prime}\right. \\
&\quad \times \left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right) \\
&\quad \times \left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right) v_{lv}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} \\
&\quad \left. \times z_{it}^{(\xi)}\eta_{it}\eta_{js}\left(v_{js}'\tilde{\Sigma}_\xi v_{js}\right)^{-1}\hat{r}_{js;r}\right] \\
&= O\left(N^{-2} + T^{-2}\right),
\end{aligned}$$

by (C.22), and the q,r-th element of $\mathrm{Cov}\left[T_3 T_2\right]$ is given by

$$\begin{aligned}
\mathrm{Cov}\left[T_3, T_2\right] &= (NT)^{-2} \sum_{i,j,k,l,m}^{N} \sum_{s,t,u,v,x}^{T} \mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-3} v_{it}^{(\xi)\prime}\right. \\
&\quad \times \left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right) \\
&\quad \times \left(\eta_{lv}^2 - \mathrm{E}\left[\eta_{lv}^2\right]\right) v_{lv}^{(\xi)\prime}\left[V^{(\xi)\prime}V^{(\xi)}\right]^{-1} \\
&\quad \times v_{it}^{(\xi)}\eta_{it}\eta_{js}\left(\eta_{mx}^2 - \mathrm{E}\left[\eta_{mx}^2\right]\right) v_{mx}^{(\xi)} \\
&\quad \left. \times \left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{js}^{(\xi)}\left(v_{js}'\tilde{\Sigma}_\xi v_{js}\right)^{-2}\hat{r}_{js;q}\right] \\
&\quad - (NT)^{-2}\left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} z_{it}^{(\xi)\prime}\right.\right. \\
&\quad \times \left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{js}^{(\xi)}\left(\eta_{js}^2 - \mathrm{E}\left[\eta_{js}^2\right]\right) \\
&\quad \left.\left. \times \left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right) v_{ku}^{(\xi)\prime}\left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{it}^{(\xi)}\eta_{it}\right]\right) \\
&\quad \times \left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\hat{r}_{it;q}\left(v_{it}'\tilde{\Sigma}_\xi v_{it}\right)^{-2} v_{it}^{(\xi)\prime}\right.\right. \\
&\quad \left.\left. \times \left[Z^{(\xi)\prime}Z^{(\xi)}\right]^{-1} v_{ku}^{(\xi)}\left(\eta_{ku}^2 - \mathrm{E}\left[\eta_{ku}^2\right]\right)\eta_{it}\right]\right) \\
&= O\left(N^{-3} + T^{-3}\right),
\end{aligned}$$

because

$$\begin{aligned}
&\sum_{i,j,k,l,m}^{N}\sum_{s,t,u,v,x}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\left(\eta_{lv}^{2}-\mathrm{E}\left[\eta_{lv}^{2}\right]\right)\eta_{it}\eta_{js}\left(\eta_{mx}^{2}-\mathrm{E}\left[\eta_{mx}^{2}\right]\right)\right]\\
&-\left(\sum_{i,j,k}^{N}\sum_{s,t,u}^{T}\mathrm{E}\left[\left(\eta_{js}^{2}-\mathrm{E}\left[\eta_{js}^{2}\right]\right)\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\right]\right)\left(\sum_{i,k}^{N}\sum_{t,u}^{T}\mathrm{E}\left[\left(\eta_{ku}^{2}-\mathrm{E}\left[\eta_{ku}^{2}\right]\right)\eta_{it}\right]\right)\\
=\;& O\left(N^{2}T^{5}+N^{5}T^{2}\right).
\end{aligned}$$

Similarly, the following equalities can be shown: $\mathrm{Var}\left[T_j\right] = O\left(N^{-j}+T^{-j}\right), \forall j \geq 4$, $\mathrm{Cov}\left[T_i, T_j\right] = O_p\left(N^{-\lceil (i+j)/2 \rceil}+T^{-\lceil (i+j)/2 \rceil}\right), \forall i,j \geq 4$, where $\lceil x \rceil$ is the smallest integer greater than or equal to x.

It also holds that $(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(B)}\mathrm{Diag}_{i=1}^{N}\left[\hat{E}_i^{-1}\right]V^{(B)\prime}\hat{D}^{-1}H$ has as q^{th} element

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{r}_{it;q}\left(v_{it}'\hat{\tilde{\Sigma}}_{\xi}v_{it}\right)^{-1}v_{it}'\left[\sum_{r=1}^{T}v_{ir}\left(v_{ir}'\hat{\tilde{\Sigma}}_{\xi}v_{ir}\right)^{-1}v_{ir}'\right]^{-1}v_{is}\left(v_{is}'\hat{\tilde{\Sigma}}_{\xi}v_{is}\right)^{-1}\eta_{is},$$

$(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(T)}\hat{G}_1^{-1}V^{(T)\prime}\hat{D}^{-1}H$ has as q^{th} element

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{r}_{it;q}\left(v_{it}'\hat{\tilde{\Sigma}}_{\xi}v_{it}\right)^{-1}v_{it}'\left[\sum_{h=1}^{N}v_{ht}\left(v_{ht}'\hat{\tilde{\Sigma}}_{\xi}v_{ht}\right)^{-1}v_{ht}'\right]^{-1}v_{jt}\left(v_{jt}'\hat{\tilde{\Sigma}}_{\xi}v_{jt}\right)^{-1}\eta_{jt}$$

and $(NT)^{-1}\hat{R}'\hat{D}^{-1}V^{(T)}\hat{G}_1^{-1}\hat{H}_1\hat{G}_1^{-1}V^{(T)\prime}\hat{D}^{-1}H$ has as q^{th} element

$$\begin{aligned}
&(NT)^{-1}\sum_{i,j}^{N}\sum_{s,t}^{T}\left\{\hat{r}_{it;q}\left(v_{it}'\hat{\tilde{\Sigma}}_{\xi}v_{it}\right)^{-1}v_{it}'\left[\sum_{h=1}^{N}v_{ht}\left(v_{ht}'\hat{\tilde{\Sigma}}_{\xi}v_{ht}\right)^{-1}v_{ht}'\right]^{-1}\right.\\
&\left[\sum_{h=1}^{N}v_{ht}\left(v_{ht}'\hat{\tilde{\Sigma}}_{\xi}v_{ht}\right)^{-1}v_{ht}'\left[\sum_{r=1}^{T}v_{hr}\left(v_{hr}'\hat{\tilde{\Sigma}}_{\xi}v_{hr}\right)^{-1}v_{hr}'\right]^{-1}v_{hs}\left(v_{hs}'\hat{\tilde{\Sigma}}_{\xi}v_{hs}\right)^{-1}v_{hs}'\right]\\
&\left.\left[\sum_{h=1}^{N}v_{hs}\left(v_{hs}'\hat{\tilde{\Sigma}}_{\xi}v_{hs}\right)^{-1}v_{hs}'\right]^{-1}v_{js}\left(v_{js}'\hat{\tilde{\Sigma}}_{\xi}v_{js}\right)^{-1}\eta_{js}\right\}
\end{aligned}$$

From which the remark below Lemma C.1 can be verified, i.e. the terms of $\hat{R}'\Phi_j Y$, $j = B, C, D$, converge at the same rates as the corresponding terms of $\hat{R}'\hat{D}^{-1}Y$.

The variance of $(NT)^{-1}\hat{R}'\hat{D}^{-1}H$ is now equal to $O\left((NT)^{-1}+N^{-2}+T^{-2}\right)$ and it holds

that $(NT)^{-1}\hat{R}'\hat{D}^{-1}H = o_p(1)$, by the Chebyshev inequality. Suppose that $N \sim T^{\kappa}$, then, since $(NT)^{-1}\hat{R}'\hat{D}^{-1}R = O_p(1)$, it holds that

$$\text{Var}\left[\hat{\varphi}_{F2SA} - \varphi\right] = \begin{cases} O\left(N^{-2}\right) \text{ if } \kappa < 1 \\ O\left(N^{-1}T^{-1}\right), \text{ if } \kappa = 1 \\ O\left(T^{-2}\right), \text{ if } \kappa > 1 \end{cases},$$

which can be summarized as

$$\min(N,T)\left(\hat{\varphi}_{F2SA} - \varphi\right) = O_p(1),$$

which proves part 2 of the theorem. ■

Remark. Note that in the second case it does not hold that

$$\text{plim}_{N,T\to\infty}\left[(NT)^{-1}\Psi_{\Omega}\hat{R}'\left(\hat{\Omega}_{\eta}^{-1} - \Omega_{\eta}^{-1}\right)H\right] = 0,$$

and thus the asymptotic equivalence between Aitken and feasible Aitken estimator can not be proved.

Appendix D

Unit Roots in Panels: Proofs of Theorems

D.1 Preliminaries

Under the unit root hypothesis, we have, from (5.1)-(5.3),

$$\begin{aligned}
\Delta y_{it} &= \tau_i + u_{it} = \tau_i + \rho u_{i,t-1} + \varepsilon_{it} = \tau_i + \sum_{s=S}^{t} \rho^{t-s} \varepsilon_{is}, \\
y_{it} &= y_{i,0} + t \cdot \tau_i + \sum_{s=1}^{t} \sum_{r=S}^{s} \rho^{s-r} \varepsilon_{ir} \\
&= y_{i,0} + t \cdot \tau_i + \sum_{s=1}^{t} \frac{1-\rho^{t-s+1}}{1-\rho} \varepsilon_{is} + \frac{1-\rho^{t}}{1-\rho} \sum_{s=S}^{0} \rho^{1-s} \varepsilon_{is},
\end{aligned}$$

where $y_{i0} = \tau_i = 0$ for Model 1 and $\tau_i = 0$ for Model 2. Define the $T \times T$ matrices

$$K_A = (k_A)_{st} = \begin{cases} 0, & s < t, \\ \frac{1-\rho^{s-t+1}}{1-\rho} & s \geq t, \end{cases}$$

and

$$L_A = (l_A)_{st} = \begin{cases} 0, & s < t, \\ \rho^{s-t}, & s \geq t. \end{cases}$$

Let $\varepsilon_i^A = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, $\mathbf{y}_{i0} = y_{i0}\iota_T$, $\mathbf{s}_i = \tau_i\iota_T$, $\mathbf{t}_i = \tau_i\kappa_T$, with ι_T a $T \times 1$ vector of ones and $\kappa_T' = (1, 2, \ldots, T)$. Furthermore, define the selection matrices

$$S_{T-j} = \begin{pmatrix} O_{(T-j)\times j} & I_{T-j} \end{pmatrix} = (s_{T-j})_{st} = \begin{cases} 1, & s = t-j, \\ 0, & \text{otherwise}, \end{cases} \tag{D.1}$$

$$T_{T-j} = \begin{pmatrix} O_{(T-j)\times(j-1)} & I_{T-j} & O_{(T-j)\times 1} \end{pmatrix} = (t_{T-j})_{st} = \begin{cases} 1, & s = t-j+1, \\ 0, & \text{otherwise}, \end{cases} \tag{D.2}$$

where $j = 1$ for Model 1 and $j = 2$ for Models 2 and 3, and where $O_{M\times N}$ is a $M \times N$ matrix of zeroes. For $S \leq 0$, define the $T \times (|S| + 1)$ matrices

$$K_B = (k_B)_{st} = \frac{1-\rho^s}{1-\rho}\rho^{2-t-S}$$

and

$$L_B = (l_B)_{st} = \rho^{1+s-t-S},$$

and the $(|S| + 1) \times 1$ vector $\varepsilon_i^B = (\varepsilon_{i,S}, \ldots, \varepsilon_{i0})'$. For $S = 1$, take K_B, L_B and ε_i^B to be empty. With the above definitions, the vectors defined in (5.9) can be written as

$$\begin{aligned} \varepsilon_i &= S_{T-j}\varepsilon_i^A, \\ y_{i,-1} &= T_{T-j}\left(\mathbf{y}_{i0} + \mathbf{t}_i + K_A\varepsilon_i^A + K_B\varepsilon_i^B\right), \\ \Delta y_{i,-1} &= T_{T-j}\left(\mathbf{s}_i + L_A\varepsilon_i^A + L_B\varepsilon_i^B\right), \end{aligned}$$

Application of the appropriate transformation (5.10)-(5.12) results in

$$\begin{aligned} Q_{T-j}\varepsilon_i &= Q_{T-j}S_{T-j}\varepsilon_i^A, \\ Q_{T-j}y_{i,-1} &= Q_{T-j}T_{T-j}K_A\varepsilon_i^A + Q_{T-j}T_{T-j}K_B\varepsilon_i^B, \\ Q_{T-j}\Delta y_{i,-1} &= Q_{T-j}T_{T-j}L_A\varepsilon_i^A + Q_{T-j}T_{T-j}L_B\varepsilon_i^B. \end{aligned}$$

Now, it follows from (5.13), (5.15) and the independence of ε_i^A and ε_i^B that

$$\begin{aligned} P &= \sigma_\varepsilon^{-2}\mathrm{E}\left[y_{i,-j}'Q_{T-j}y_{i,-j}\right] \\ &= \sigma_\varepsilon^{-2}\mathrm{E}\left[\varepsilon_i^{A\prime}K_A'T_{T-j}'Q_{T-j}T_{T-j}K_A\varepsilon_i^A\right] + \sigma_\varepsilon^{-2}\mathrm{E}\left[\varepsilon_i^{B\prime}K_B'T_{T-j}'Q_{T-j}T_{T-j}K_B\varepsilon_i^B\right] \\ &= \mathrm{Tr}\left(T_{T-j}'Q_{T-j}T_{T-j}K_AK_A'\right) + \mathrm{Tr}\left(T_{T-j}'Q_{T-j}T_{T-j}K_BK_B'\right). \end{aligned} \tag{D.3}$$

Similarly,

$$\begin{aligned} Q &= \sigma_\varepsilon^{-2}\mathrm{E}\left[\Delta y'_{i,-j}Q_{T-j}y_{i,-j}\right] \\ &= \mathrm{Tr}\left(T'_{T-j}Q_{T-j}T_{T-j}L_AK'_A\right)+\mathrm{Tr}\left(T'_{T-j}Q_{T-j}T_{T-j}L_BK'_B\right), \end{aligned} \tag{D.4}$$

$$\begin{aligned} R &= \sigma_\varepsilon^{-2}\mathrm{E}\left[\Delta y'_{i,-j}Q_{T-j}\Delta y_{i,-j}\right] \\ &= \mathrm{Tr}\left(T'_{T-j}Q_{T-j}T_{T-j}L_AL'_A\right)+\mathrm{Tr}\left(T'_{T-j}Q_{T-j}T_{T-j}L_BL'_B\right), \end{aligned} \tag{D.5}$$

$$V = \sigma_\varepsilon^{-2}\mathrm{E}\left[\varepsilon'_iQ_{T-j}y_{i,-j}\right] = \mathrm{Tr}\left(S'_{T-j}Q_{T-j}T_{T-j}K_A\right), \tag{D.6}$$

$$W = \sigma_\varepsilon^{-2}\mathrm{Tr}\left[\varepsilon'_iQ_{T-j}\Delta y_{i,-j}\right] = \mathrm{Tr}\left(S'_{T-j}Q_{T-j}T_{T-j}L_A\right). \tag{D.7}$$

In the proofs the following matrix products are needed

$$K_AK'_A = (kk_A)_{st} = \begin{cases} \sum_{r=1}^{s}\frac{(1-\rho^r)\left(1-\rho^{r+t-s}\right)}{(1-\rho)^2}, & s<t, \\ \sum_{r=1}^{t}\frac{\left(1-\rho^{r+s-t}\right)(1-\rho^r)}{(1-\rho)^2}, & s\geq t, \end{cases} \tag{D.8}$$

$$L_AK'_A = (lk_A)_{st} = \begin{cases} \sum_{r=1}^{s}\frac{\left(1-\rho^{r+t-s}\right)\rho^{r-1}}{1-\rho}, & s<t, \\ \sum_{r=1}^{t}\frac{(1-\rho^r)\rho^{r+s-t-1}}{1-\rho}, & s\geq t, \end{cases} \tag{D.9}$$

$$L_AL'_A = (ll_A)_{st} = \begin{cases} \sum_{r=1}^{s}\rho^{2r+t-s-2}, & s<t, \\ \sum_{r=1}^{t}\rho^{2r+s-t-2}, & s\geq t, \end{cases} \tag{D.10}$$

$$K_BK'_B = (kk_B)_{st} = \omega\frac{1-\rho^s}{1-\rho}\frac{1-\rho^t}{1-\rho}, \tag{D.11}$$

$$L_BK'_B = (lk_B)_{st} = \omega\rho^{s-1}\frac{1-\rho^t}{1-\rho}, \tag{D.12}$$

$$L_BL'_B = (ll_B)_{st} = \omega\rho^{s+t-2}, \tag{D.13}$$

with

$$\omega = \sum_{r=S}^{0}\rho^{2-2r},$$

and, for any $(T-j)\times(T-j)$ matrix X,

$$T'_{T-j}XT_{T-j} = (txt_{T-j})_{st} = \begin{cases} x_{s-j+1,t-j+1},\ s,t=j,\ldots,T-1, \\ 0, \qquad \text{otherwise}, \end{cases} \tag{D.14}$$

$$S'_{T-j}XT_{T-j} = (sxt_{T-j})_{st} = \begin{cases} x_{s-j+1,t-j},\ s=j,\ldots,T-1; t=j+1,\ldots,T, \\ 0, \qquad \text{otherwise}. \end{cases} \tag{D.15}$$

D.2 Proofs for Model 1

Proof of Theorem 5.1. Recalling, (5.14)-(5.15), the proof is complete if it is shown that P, Q and R are as stated in the Theorem, that $V=W=0$ and that

$$\text{Var}\sqrt{N}\begin{pmatrix} \bar{v} \\ \bar{w} \end{pmatrix} = \begin{pmatrix} P & Q \\ Q & R \end{pmatrix}.$$

We have $j=1$ and $Q_{T-j}=I_{T-1}$, so $T'_{T-j}Q_{T-j}T_{T-j}=T'_{T-1}T_{T-1}$, which is given by

$$T'_{T-1}T_{T-1} = (tt_{T-1})_{st} = \begin{cases} 1,\ s=t=1,\ldots,T-1, \\ 0,\ \text{otherwise}. \end{cases}$$

This is a $(T-1)\times(T-1)$ identity matrix expanded with one column of zeroes to the right, and one bottom row of zeroes. Hence, from (D.3), P equals the sum of the first $T-1$ diagonal elements of $K_AK'_A+K_BK'_B$, and similarly for Q and R. From (D.8)-(D.13), we have

$$\begin{array}{ll} (kk_A)_{tt}=\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2, & (kk_B)_{tt}=\omega\left(\frac{1-\rho^t}{1-\rho}\right)^2, \\ (lk_A)_{tt}=\sum_{s=1}^{t}\rho^{s-1}\frac{1-\rho^s}{1-\rho}, & (lk_B)_{tt}=\omega\rho^{t-1}\frac{1-\rho^t}{1-\rho}, \\ (ll_A)_{tt}=\sum_{s=1}^{t}\rho^{2s-2}, & (ll_B)_{tt}=\omega\rho^{2t-2}. \end{array}$$

The stated expressions for P, Q and R follow. Using (D.15), it holds that

$$S'_{T-1}T_{T-1} = (st_{T-1})_{st} = \begin{cases} 1, & s=t+1=2,\ldots,T, \\ 0, & \text{otherwise} \end{cases}$$

Since $S'_{T-1}T_{T-1}$ is lower diagonal with diagonal elements equal to zero, and K_A is lower diagonal, $S'_{T-1}T_{T-1}K_A$ has a trace equal to zero, and similarly for $S'_{T-1}T_{T-1}L_A$. Hence, from

(D.6)-(D.7), $V = W = 0$. Furthermore, by the independence of ε_{it} and $(y_{i,t-1}, \Delta y_{i,t-1})$,

$$\begin{aligned}
\mathrm{Var}\left[\sqrt{N}\overline{v}\right] &= \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} y_{i,t-1}\varepsilon_{it}\right)^2\right] = \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} y_{i,t-1}^2\varepsilon_{it}^2\right)\right] = P, \\
\mathrm{Var}\left[\sqrt{N}\overline{w}\right] &= \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} \Delta y_{i,t-1}\varepsilon_{it}\right)^2\right] = \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} \left(\Delta y_{i,t-1}\right)^2\varepsilon_{it}^2\right)\right] = R
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}\left[\sqrt{N}\overline{v}, \sqrt{N}\overline{w}\right] &= \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} y_{i,t-1}\varepsilon_{it}\right)\left(\sum_{t=2}^{T} \Delta y_{i,t-1}\varepsilon_{it}\right)\right] \\
&= \sigma_{\varepsilon}^{-4}\mathrm{E}\left[\left(\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t-1}\varepsilon_{it}^2\right)\right] = Q.
\end{aligned}$$

This completes the proof. ■

Remark. The probability limits P, Q and R appearing in Theorem 5.1 can further be rewritten as

$$\begin{aligned}
P &= \frac{T(T-1)}{2(1-\rho)^2} - \frac{(T-1)(2\rho+\rho^2)}{(1-\rho)^2(1-\rho^2)} + 2\frac{\rho^2-\rho^{T+1}}{(1-\rho)^4} - \frac{\rho^4-\rho^{2T+2}}{(1-\rho)^2(1-\rho^2)^2} \\
&\quad +\omega\frac{(T-1)}{(1-\rho)^2} - 2\omega\frac{\left(\rho-\rho^T\right)}{(1-\rho)^3} + \omega\frac{\left(\rho^2-\rho^{2T}\right)}{(1-\rho)^2(1-\rho^2)}, \\
Q &= \frac{(T-1)}{(1-\rho)(1-\rho^2)} - \frac{\rho-\rho^T}{(1-\rho)^3} + \frac{\rho^3-\rho^{2T+1}}{(1-\rho)(1-\rho^2)^2} \\
&\quad +\omega\frac{\left(1-\rho^{T-1}\right)}{(1-\rho)^2} - \omega\frac{\left(\rho-\rho^{2T-1}\right)}{(1-\rho)(1-\rho^2)}, \\
R &= \frac{(T-1)}{1-\rho^2} - \frac{\rho^2-\rho^{2T}}{(1-\rho^2)^2} + \omega\frac{\left(1-\rho^{2T-2}\right)}{(1-\rho^2)}.
\end{aligned} \tag{D.16}$$

Proof of Lemma 5.1. $S = 1$ implies $\omega = 0$, so we can rewrite $P(\rho, T)$ as

$$P(\rho, T) = \sum_{t=2}^{T}\sum_{s=1}^{t-1}\left(\frac{1-\rho^{t-s}}{1-\rho}\right)^2 = \sum_{t=2}^{T}\sum_{p=0}^{t-2}\left(\sum_{q=0}^{p}\rho^q\right)^2$$

and $Q(\rho, T)$ as

$$Q(\rho, T) = \sum_{t=2}^{T} \sum_{s=1}^{t-1} \rho^{t-1-s} \frac{1-\rho^{t-s}}{1-\rho} = \sum_{t=2}^{T} \sum_{p=0}^{t-2} \rho^{p} \sum_{q=0}^{p} \rho^{q},$$

using

$$\frac{1-\rho^{t-s}}{1-\rho} = \sum_{r=s}^{t-1} \rho^{r-s}$$

and the substitutions $q = r - s$ and $p = t - s - 1$. An induction on $p \geq 0$ shows that

$$\begin{array}{lll} 0 < \sum_{q=0}^{p} \rho^{q} \leq 1, & \text{if} & -1 < \rho < 0; \\ 1 = \sum_{q=0}^{p} \rho^{q}, & \text{if} & \rho = 0; \\ 1 \leq \sum_{q=0}^{p} \rho^{q}, & \text{if} & 0 < \rho < 1; \end{array} \tag{D.17}$$

with strict inequalities for $p \geq 1$. From expression (D.17) the inequalities (5.17) follow immediately, since $T \geq 3$ and

$$\sum_{t=2}^{T} \left(\sum_{p=0}^{t-2} 1 \right) = \frac{T(T-1)}{2}.$$

An induction on $t \geq 2$, using (D.17) shows that

$$\sum_{p=0}^{t-2} \rho^{p} \sum_{q=0}^{p} \rho^{q} \geq 0, \forall \rho \in [-1, \infty),$$

with strict inequality for $t > 2$. Hence $Q(\rho, T) > 0$. ■

D.3 Proofs for Model 2

The proofs for Models 2 and 3 will make use of the following lemma.

Lemma D.1. *For arbitrary functions $f(\cdot)$ and $g(\cdot)$, the following identities hold:*

$$\begin{aligned}
&\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1}\sum_{s=t}^{T-1}\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\sum_{t=2}^{T-1}\sum_{s=1}^{t} f(s)\, g(s) \\
&+ \sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s} \left(f(r)\, g(r+t-s) + f(r+t-s)\, g(r)\right), \qquad \text{(D.18)}
\end{aligned}$$

$$\begin{aligned}
&\sum_{t=2}^{T-1} t \sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1} t \sum_{s=t}^{T-1}\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
&+ \sum_{t=2}^{T-1}\sum_{s=2}^{t-1} s \sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1}\sum_{s=t}^{T-1} s \sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\; 2\sum_{t=2}^{T-1} t \sum_{s=1}^{t} f(s)\, g(s) \\
&+ \sum_{t=2}^{T-1}\sum_{s=2}^{t-1} (t+s) \sum_{r=1}^{s} \left(f(r)\, g(r+t-s) + f(r+t-s)\, g(r)\right), \qquad \text{(D.19)}
\end{aligned}$$

$$\begin{aligned}
&\sum_{t=2}^{T-1} t \sum_{s=2}^{t-1} s \sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1} t \sum_{s=t}^{T-1} s \sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\sum_{t=2}^{T-1} t^2 \sum_{s=1}^{t} f(s)\, g(s) \\
&+ \sum_{t=2}^{T-1} t \sum_{s=2}^{t-1} s \sum_{r=1}^{s} \left(f(r)\, g(r+t-s) + g(r)\, f(r+t-s)\right). \qquad \text{(D.20)}
\end{aligned}$$

Proof. The left-hand side of (D.18) can be written as

$$\begin{aligned}
&\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) \\
&+\sum_{t=2}^{T-1}\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1}\sum_{s=t+1}^{T-1}\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\sum_{t=2}^{T-1}\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{s=2}^{T-1}\sum_{t=2}^{s-1}\sum_{r=1}^{t} f(r+s-t)\, g(r),
\end{aligned}$$

whence (D.18). The left-hand side of (D.19) can be written as

$$\begin{aligned}
&\sum_{t=2}^{T-1} t\sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1} t\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1} t\sum_{s=t+1}^{T-1}\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
&+\sum_{t=2}^{T-1}\sum_{s=2}^{t-1} s\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1} t\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1}\sum_{s=t+1}^{T-1} s\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\; 2\sum_{t=2}^{T-1} t\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1} t\sum_{s=2}^{t-1}\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{s=2}^{T-1}\sum_{t=2}^{s-1} t\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
&+\sum_{t=2}^{T-1}\sum_{s=2}^{t-1} s\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{s=2}^{T-1} s\sum_{t=2}^{s-1}\sum_{r=1}^{t} f(r+s-t)\, g(r).
\end{aligned}$$

Collecting terms gives (D.19). The left-hand side of (D.20) can be written as

$$\begin{aligned}
&\sum_{t=2}^{T-1} t\sum_{s=2}^{t-1} s\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{t=2}^{T-1} t^2\sum_{r=1}^{t} f(r)\, g(r) + \sum_{t=2}^{T-1} t\sum_{s=t+1}^{T-1} s\sum_{r=1}^{t} f(r+s-t)\, g(r) \\
= &\sum_{t=2}^{T-1} t^2\sum_{r=1}^{t} g(r)\, f(r) + \sum_{t=2}^{T-1} t\sum_{s=2}^{t-1} s\sum_{r=1}^{s} f(r)\, g(r+t-s) + \sum_{s=2}^{T-1} s\sum_{t=2}^{s-1} t\sum_{r=1}^{t} f(r+s-t)\, g(r),
\end{aligned}$$

whence (D.20). ■

Proof of Theorem 5.2. Under the assumptions made and given that $Q_{T-j} = I_{T-2} - \frac{\iota_{T-2}\iota'_{T-2}}{T-2}$, we obtain

$$P = \operatorname{Tr}\left(T'_{T-2}Q_{T-2}T_{T-2}K_AK'_A\right) + \operatorname{Tr}\left(T'_{T-2}Q_{T-2}T_{T-2}K_BK'_B\right),$$

which can be rewritten as

$$\begin{aligned} P &= \mathrm{Tr}\left(T'_{T-2}T_{T-2}K_AK'_A\right) - \frac{1}{T-2}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_AK'_A\right) \\ &\quad +\mathrm{Tr}\left(T'_{T-2}T_{T-2}K_BK'_B\right) - \frac{1}{T-2}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_BK'_B\right). \end{aligned}$$

Taking into account the matrix products (D.8), (D.11) and (D.14), it holds that

$$\begin{aligned} T'_{T-2}T_{T-2} &= (tt_{T-2})_{st} = \begin{cases} 1, & s=t=2,\ldots,T-1, \\ 0, & \text{otherwise}, \end{cases} \\ T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2} &= (tiit_{T-2})_{st} = \begin{cases} 1, & s,t=2,\ldots,T-1, \\ 0, & \text{otherwise}. \end{cases} \end{aligned}$$

The tth diagonal element of $T'_{T-2}T_{T-2}K_AK'_A$ and $T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_AK'_A$ are equal to

$$\sum_{q=1}^{T}(tt_{T-2})_{tq}(kk_A)_{qt} = (kk_A)_{tt} = \sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2$$

and

$$\sum_{q=1}^{T}(tiit_{T-2})_{tq}(kk_A)_{qt} = \sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho} + \sum_{s=t}^{T-1}\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho},$$

respectively, for $2 \leq t \leq T-1$, and the tth diagonal element of $T'_{T-1}T_{T-1}K_BK'_B$ and $T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_BK'_B$ are given by

$$\sum_{q=1}^{T}(tt_{T-2})_{tq}(kk_B)_{qt} = (kk_B)_{tt} = \omega\left(\frac{1-\rho^t}{1-\rho}\right)^2$$

and

$$\sum_{q=1}^{T}(tiit_{T-2})_{tq}(kk_B)_{qt} = \omega\frac{1-\rho^t}{1-\rho}\left(\sum_{s=2}^{T-1}\frac{1-\rho^s}{1-\rho}\right),$$

respectively, for $2 \leq t \leq T-1$.

Consequently,

$$\begin{aligned} P &= \sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2 - \frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho} \\ &\quad -\frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=t}^{T-1}\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho} \\ &\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2 - \frac{\omega}{T-2}\left(\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\right)^2. \end{aligned}$$

Using Lemma D.1, identity (D.18), with $f(r) = g(r) = \frac{1-\rho^r}{1-\rho}$, it holds that

$$\begin{aligned} P &= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2 - \frac{2}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{(1-\rho^r)(1-\rho^{r+t-s})}{(1-\rho)^2} \\ &\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2 - \frac{\omega}{T-2}\left(\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\right)^2. \end{aligned}$$

Similarly, by (D.19) with $f(r) = \rho^{r-1}$ and $g(r) = \frac{1-\rho^r}{1-\rho}$, it holds that

$$\begin{aligned} Q &= \mathrm{Tr}\left(T'_{T-2}T_{T-2}L_AK'_A\right) - \frac{1}{T-2}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\cdot\iota'_{T-2}\right)T_{T-2}L_AK'_A\right) \\ &\quad +\mathrm{Tr}\left(T'_{T-2}T_{T-2}L_BK'_B\right) - \frac{1}{T-2}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\cdot\iota'_{T-2}\right)T_{T-2}L_BK'_B\right) \\ &= \sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{s-1}\frac{1-\rho^s}{1-\rho} - \frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{r-1}\frac{1-\rho^{r+t-s}}{1-\rho} \\ &\quad -\frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{r+s-t-1}\frac{1-\rho^r}{1-\rho} \\ &\quad +\omega\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\rho^{t-1} - \frac{\omega}{T-2}\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1} \\ &= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\frac{\rho^{s-1}(1-\rho^s)}{1-\rho} \\ &\quad -\frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{\rho^{r-1}(1-\rho^{r+t-s}) + \rho^{r+t-s-1}(1-\rho^r)}{1-\rho} \\ &\quad +\omega\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\rho^{t-1} - \frac{\omega}{T-2}\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}, \end{aligned}$$

and, by (D.20) with $f(r) = g(r) = \rho^{r-1}$,

$$\begin{aligned}
R &= \operatorname{Tr}\left(T'_{T-2}T_{T-2}L_AL'_A\right) - \frac{1}{T-2}\operatorname{Tr}\left(T'_{T-2}\left(\iota_{T-2}\cdot\iota'_{T-2}\right)T_{T-2}L_AL'_A\right) \\
&\quad + \operatorname{Tr}\left(T'_{T-2}T_{T-2}L_BL'_B\right) - \frac{1}{T-2}\operatorname{Tr}\left(T'_{T-2}\left(\iota_{T-2}\cdot\iota'_{T-2}\right)T_{T-2}L_BL'_B\right) \\
&= \sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2} - \frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{r-1}\rho^{r+t-s-1} - \frac{1}{T-2}\sum_{t=2}^{T-1}\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{r+s-t-1}\rho^{r-1} \\
&\quad + \sum_{t=2}^{T-1}\rho^{2t-2}\sum_{s=S}^{0}\rho^{2-2s} - \frac{1}{T-2}\sum_{t=2}^{T-1}\rho^{t-1}\sum_{s=2}^{T-1}\rho^{s-1}\sum_{r=S}^{0}\rho^{2-2r} \\
&= \frac{T-3}{T-2}\sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2} - \frac{2}{T-2}\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2} + \omega\sum_{t=2}^{T-1}\rho^{2t-2} - \frac{\omega}{T-2}\left(\sum_{t=2}^{T-1}\rho^{t-1}\right)^2.
\end{aligned}$$

Taking into account the matrix products (D.15), it holds that

$$\begin{aligned}
S'_{T-2}T_{T-2} &= \left(ts_{T-2}\right)_{st} = \begin{cases} 1, \ t = s+1 = 3,\ldots,T, \\ 0, \ \text{otherwise}, \end{cases} \\
S'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2} &= \left(tiis_{T-2}\right)_{st} = \begin{cases} 1, \ s = 2,\ldots,T-1, \text{ and } t = 3,\ldots,T, \\ 0, \ \text{otherwise}, \end{cases}
\end{aligned}$$

and the diagonal elements of $S'_{T-2}T_{T-2}K_A$ are equal to zero and the tth diagonal element of $S'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_A$ is given by

$$\sum_{q=1}^{T}\left(siit_{T-2}\right)_{tq}(k)_{tq} = \sum_{r=3}^{t}\frac{1-\rho^{t-r+1}}{1-\rho},$$

for $3 \leq t \leq T-1$. Therefore,

$$V = \operatorname{Tr}\left(S'_{T-2}T_{T-2}K_A\right) - \frac{1}{T-2}\operatorname{Tr}\left(S'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_A\right) = -\frac{1}{T-2}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho},$$

and similarly,

$$W = \operatorname{Tr}\left(S'_{T-2}T_{T-2}L_A\right) - \frac{1}{T-2}\operatorname{Tr}\left(S'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}L_A\right) = -\frac{1}{T-2}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\rho^{t-s}.$$

Application of the central limit theorem and Slutzky's theorem gives the desired result. ■

D.4 Proof for Model 3

Proof of Theorem 5.3. Let

$$\begin{aligned}
S_0 &= \sum_{s=1}^{T-2} 1 = (T-2), \\
S_1 &= \sum_{s=1}^{T-2} s = \frac{(T-1)(T-2)}{2}, \\
S_2 &= \sum_{s=1}^{T-2} s^2 = \frac{(T-1)(T-2)(2T-3)}{6}, \\
S_3 &= \det\left[Z'_{T-2}Z_{T-2}\right] = \left(\sum_{s=1}^{T-2} s^2\right)\left(\sum_{s=1}^{T-2} 1\right) - \left(\sum_{s=1}^{T-2} s\right)^2 = \frac{(T-3)(T-2)^2(T-1)}{12},
\end{aligned}$$

and

$$\begin{aligned}
A &= \frac{S_2 + 3S_1 + 2S_0}{S_3} = \frac{4(T^2+2T+3)}{(T-1)(T-2)(T-3)}, \\
B &= \frac{S_1 + 2S_0}{S_3} = \frac{6(T+3)}{(T-1)(T-2)(T-3)}, \\
C &= \frac{S_1 + S_0}{S_3} = \frac{6(T+1)}{(T-1)(T-2)(T-3)}, \\
D &= \frac{S_0}{S_3} = \frac{12}{(T-1)(T-2)(T-3)}, \\
E &= \frac{S_3 - S_2 - 2S_1 - S_0}{S_3} = \frac{T^3 - 10T^2 + 9T - 12}{(T-1)(T-2)(T-3)}, \\
F &= \frac{S_2 + 2S_1 + S_0}{S_3} = \frac{2(2T^2+T+3)}{(T-1)(T-2)(T-3)}.
\end{aligned}$$

Then, since $Q_{T-2} = I_{T-2} - Z_{T-2}\left(Z'_{T-2}Z_{T-2}\right)^{-1}Z'_{T-2}$, we can write

$$Q_{T-2} = I_{T-2} - \frac{S_2\iota_{T-2}\iota'_{T-2} - S_1\left(\kappa_{T-2}\iota'_{T-2} + \iota_{T-2}\kappa'_{T-2}\right) + S_0\left(\kappa_{T-2}\kappa'_{T-2}\right)}{S_3},$$

and

$$\begin{aligned}
P &= \mathrm{Tr}\left(T'_{T-2}T_{T-2}K_AK'_A\right) - \frac{S_2}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_AK'_A\right) \\
&\quad + \frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}K_AK'_A\right) + \frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}K_AK'_A\right)
\end{aligned}$$

$$
\begin{aligned}
\ldots \quad & -\frac{S_0}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}K_AK'_A\right)+\mathrm{Tr}\left(T'_{T-2}T_{T-2}K_BK'_B\right)\\
& -\frac{S_2}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}K_BK'_B\right)+\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}K_BK'_B\right)\\
& +\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}K_BK'_B\right)-\frac{S_0}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}K_BK'_B\right).
\end{aligned}
$$

Taking into account the matrix products (D.8), (D.11) and (D.14), it holds that

$$
\begin{aligned}
T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2} &= \left(tkit_{T-2}\right)_{st}=\begin{cases} s-1,\ s,t=2,\ldots,T-1,\\ 0,\quad \text{otherwise,}\end{cases}\\
T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2} &= \left(tkkt_{T-2}\right)_{st}=\begin{cases} (s-1)(t-1),\ s,t=2,\ldots,T-1,\\ 0,\quad \text{otherwise.}\end{cases}
\end{aligned}
$$

Therefore, the tth diagonal element of $T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}K_AK'_A$, $T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}K_AK'_A$ and $T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}K_AK'_A$ are equal to

$$
\begin{aligned}
\sum_{s=1}^{T}\left(tkit_{T-2}\right)_{ts}\left(kk_A\right)_{st} &= (t-1)\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+(t-1)\sum_{s=t}^{T-1}\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho},\\
\sum_{s=1}^{T}\left(tikt_{T-2}\right)_{ts}\left(kk_A\right)_{st} &= \sum_{s=2}^{t-1}(s-1)\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}(s-1)\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho}
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{s=1}^{T}\left(tkk_At_{T-2}\right)_{ts}\left(kk_A\right)_{st} &= (t-1)\sum_{s=2}^{t-1}(s-1)\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}\\
&\quad +(t-1)\sum_{s=t}^{T-1}(s-1)\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho},
\end{aligned}
$$

respectively, for $2\leq t\leq T-1$. Moreover, the tth diagonal element of $T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}K_BK'_B$, $T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}K_BK'_B$ and $T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}K_BK'_B$ are equal to

$$
\begin{aligned}
\sum_{s=1}^{T}\left(tkit_{T-2}\right)_{ts}\left(kk_B\right)_{st} &= \omega\,(t-1)\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\frac{1-\rho^s}{1-\rho},\\
\sum_{s=1}^{T}\left(tikt_{T-2}\right)_{ts}\left(kk_B\right)_{st} &= \omega\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}(s-1)\frac{1-\rho^s}{1-\rho},
\end{aligned}
$$

and

$$\sum_{s=1}^{T}\left(tkk_Bt_{T-2}\right)_{ts}\left(kk_B\right)_{st}=\omega\left(t-1\right)\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\left(s-1\right)\frac{1-\rho^s}{1-\rho},$$

respectively, for $2\leq t\leq T-1$. Consequently,

$$\begin{aligned}
P &= \sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2 \\
&\quad -\frac{S_2+2S_1+S_0}{S_3}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho}\right) \\
&\quad +\frac{S_1+S_0}{S_3}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho}\right) \\
&\quad +\frac{S_1+S_0}{S_3}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho}\right) \\
&\quad -\frac{S_0}{S_3}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{1-\rho^r}{1-\rho}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\frac{1-\rho^{r+s-t}}{1-\rho}\frac{1-\rho^r}{1-\rho}\right) \\
&\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2-\omega\frac{S_0+2S_1+S_2}{S_3}\left(\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\right)^2 \\
&\quad +2\omega\frac{S_1+S_0}{S_3}\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\frac{1-\rho^s}{1-\rho}-\omega\frac{S_0}{S_3}\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}s\frac{1-\rho^s}{1-\rho}.
\end{aligned}$$

Using Lemma D.1, identities (D.18-D.20), with $f\left(r\right)=g\left(r\right)=\frac{1-\rho^r}{1-\rho}$, it holds that

$$\begin{aligned}
P &= E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2-2F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{\left(1-\rho^r\right)\left(1-\rho^{r+t-s}\right)}{\left(1-\rho\right)^2} \\
&\quad +2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2+2C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\left(t+s\right)\sum_{r=1}^{s}\frac{\left(1-\rho^r\right)\left(1-\rho^{r+t-s}\right)}{\left(1-\rho\right)^2} \\
&\quad -D\sum_{t=2}^{T-1}t^2\sum_{s=1}^{t}\left(\frac{1-\rho^s}{1-\rho}\right)^2-2D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{\left(1-\rho^r\right)\left(1-\rho^{r+t-s}\right)}{\left(1-\rho\right)^2} \\
&\quad +\omega\sum_{t=2}^{T-1}\left(\frac{1-\rho^t}{1-\rho}\right)^2-\omega F\left(\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\right)^2 \\
&\quad +2\omega C\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\frac{1-\rho^s}{1-\rho}-\omega D\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}s\frac{1-\rho^s}{1-\rho}.
\end{aligned}$$

Similarly, by (D.18-D.20) with $f(r)=\rho^{r-1}$ and $g(r)=\frac{1-\rho^r}{1-\rho}$, it holds that

$$\begin{aligned}
Q &= \mathrm{Tr}\left(T'_{T-2}T_{T-2}L_AK'_A\right)-\frac{S_2}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}L_AK'_A\right)\\
&+\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}L_AK'_A\right)+\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}L_AK'_A\right)\\
&-\frac{S_0}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}L_AK'_A\right)\\
&+\mathrm{Tr}\left(T'_{T-2}T_{T-2}L_BK'_B\right)-\frac{S_2}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}L_BK'_B\right)\\
&+\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}L_BK'_B\right)+\frac{S_1}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}L_BK'_B\right)\\
&-\frac{S_0}{S_3}\mathrm{Tr}\left(T'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}L_BK'_B\right)
\end{aligned}$$

$$\begin{aligned}
Q &= \sum_{t=2}^{T-1}\left(\sum_{s=1}^{t}\rho^{s-1}\frac{1-\rho^s}{1-\rho}\right)\\
&-\frac{S_2+2S_1+S_0}{S_3}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{r-1}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{r+s-t-1}\frac{1-\rho^r}{1-\rho}\right)\\
&+\frac{S_1+S_0}{S_3}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{r-1}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{r+s-t-1}\frac{1-\rho^r}{1-\rho}\right)\\
&+\frac{S_1+S_0}{S_3}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{r-1}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\rho^{r+s-t-1}\frac{1-\rho^r}{1-\rho}\right)\\
&-\frac{S_0}{S_3}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{r-1}\frac{1-\rho^{r+t-s}}{1-\rho}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\rho^{r+s-t-1}\frac{1-\rho^r}{1-\rho}\right)\\
&+\omega\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\rho^{t-1}-\frac{\omega\left(S_0+2S_1+S_2\right)}{S_3}\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}\\
&+\frac{\omega\left(S_0+S_1\right)}{S_3}\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}+\frac{\omega\left(S_0+S_1\right)}{S_3}\sum_{t=2}^{T-1}\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1}\\
&-\frac{\omega S_0}{S_3}\sum_{t=2}^{T-1}t\frac{1-\rho^t}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1}
\end{aligned}$$

$$
\begin{aligned}
Q &= E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}-F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&+2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}+C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\left(t+s\right)\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&-D\sum_{t=2}^{T-1}t^{2}\sum_{s=1}^{t}\frac{\rho^{s-1}\left(1-\rho^{s}\right)}{1-\rho}-D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\frac{\rho^{r-1}-2\rho^{2r+t-s-1}+\rho^{r+t-s-1}}{1-\rho}\\
&+\omega\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\rho^{t-1}-\omega F\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}+\omega C\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}\rho^{s-1}\\
&+\omega C\sum_{t=2}^{T-1}\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1}-\omega D\sum_{t=2}^{T-1}t\frac{1-\rho^{t}}{1-\rho}\sum_{s=2}^{T-1}s\rho^{s-1}
\end{aligned}
$$

and, by (D.18-D.20) with $f\left(r\right)=g\left(r\right)=\rho^{r-1}$,

$$
\begin{aligned}
R &= \mathrm{Tr}\left(T_{T-2}^{\prime}T_{T-2}L_{A}L_{A}^{\prime}\right)-\frac{S_{2}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\iota_{T-2}\iota_{T-2}^{\prime}\right)T_{T-2}L_{A}L_{A}^{\prime}\right)\\
&+\frac{S_{1}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\kappa_{T-2}\iota_{T-2}^{\prime}\right)T_{T-2}L_{A}L_{A}^{\prime}\right)+\frac{S_{1}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\iota_{T-2}\kappa_{T-2}^{\prime}\right)T_{T-2}L_{A}L_{A}^{\prime}\right)\\
&-\frac{S_{0}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\kappa_{T-2}\kappa_{T-2}^{\prime}\right)T_{T-2}L_{A}L_{A}^{\prime}\right)\\
&+\mathrm{Tr}\left(T_{T-2}^{\prime}T_{T-2}L_{B}L_{B}^{\prime}\right)-\frac{S_{2}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\iota_{T-2}\iota_{T-2}^{\prime}\right)T_{T-2}L_{B}L_{B}^{\prime}\right)\\
&+\frac{S_{1}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\kappa_{T-2}\iota_{T-2}^{\prime}\right)T_{T-2}L_{B}L_{B}^{\prime}\right)+\frac{S_{1}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\iota_{T-2}\kappa_{T-2}^{\prime}\right)T_{T-2}L_{B}L_{B}^{\prime}\right)\\
&-\frac{S_{0}}{S_{3}}\mathrm{Tr}\left(T_{T-2}^{\prime}\left(\kappa_{T-2}\kappa_{T-2}^{\prime}\right)T_{T-2}L_{B}L_{B}^{\prime}\right)
\end{aligned}
$$

$$
\begin{aligned}
R &= \sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2}-\frac{S_{2}+2S_{1}+S_{0}}{S_{3}}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{2r+s-t-2}\right)\\
&+\frac{S_{1}+S_{0}}{S_{3}}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2}+\sum_{s=t}^{T-1}\sum_{r=1}^{t}\rho^{2r+s-t-2}\right)\\
&+\frac{S_{1}+S_{0}}{S_{3}}\sum_{t=2}^{T-1}\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{2r+t-s-2}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\rho^{2r+s-t-2}\right)\\
&-\frac{S_{0}}{S_{3}}\sum_{t=2}^{T-1}t\left(\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{2r+t-s-2}+\sum_{s=t}^{T-1}s\sum_{r=1}^{t}\rho^{2r+s-t-2}\right)
\end{aligned}
$$

$$\begin{aligned}\ldots \quad & +\omega\sum_{t=2}^{T-1}\rho^{2t-2}-\frac{\omega\left(S_0+2S_1+S_2\right)}{S_3}\left(\sum_{t=2}^{T-1}\rho^{t-1}\right)^2\\ & +2\frac{\omega\left(S_0+S_1\right)}{S_3}\sum_{t=2}^{T-1}t\rho^{t-1}\sum_{s=2}^{T-1}\rho^{s-1}-\frac{\omega S_0}{S_3}\sum_{t=2}^{T-1}t\rho^{t-1}\sum_{s=2}^{T-1}s\rho^{s-1}\end{aligned}$$

$$\begin{aligned}R \;=\; & E\sum_{t=2}^{T-1}\sum_{s=1}^{t}\rho^{2s-2}-2F\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}\sum_{r=1}^{s}\rho^{2r+t-s-2}\\ & +2C\sum_{t=2}^{T-1}t\sum_{s=1}^{t}\rho^{2s-2}+2C\sum_{t=2}^{T-1}\sum_{s=2}^{t-1}(t+s)\sum_{r=1}^{s}\rho^{2r+t-s-2}\\ & -D\sum_{t=2}^{T-1}t^2\sum_{s=1}^{t}\rho^{2s-2}-2D\sum_{t=2}^{T-1}t\sum_{s=2}^{t-1}s\sum_{r=1}^{s}\rho^{2r+t-s-2}\\ & +\omega\sum_{t=2}^{T-1}\rho^{2t-2}-\omega F\left(\sum_{t=2}^{T-1}\rho^{t-1}\right)^2+2\omega C\sum_{t=2}^{T-1}t\rho^{t-1}\sum_{s=2}^{T-1}\rho^{s-1}-\omega D\sum_{t=2}^{T-1}t\rho^{t-1}\sum_{s=2}^{T-1}s\rho^{s-1}.\end{aligned}$$

Taking into account the matrix products (D.15), it holds that

$$\begin{aligned}S'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2} &= \left(skit_{T-2}\right)_{st}=\begin{cases}s-1, & s=2,\ldots,T-1,t=3,\ldots,T,\\ 0, & \text{otherwise,}\end{cases}\\ S'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2} &= \left(sikt_{T-2}\right)_{st}=\begin{cases}t-2, & s=2,\ldots,T-1,t=3,\ldots,T,\\ 0, & \text{otherwise,}\end{cases}\\ S'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2} &= \left(skkt_{T-2}\right)_{st}=\begin{cases}(s-1)(t-2), & s=2,\ldots,T-1,t=3,\ldots,T,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

Consequently, the tth diagonal element of $S'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}K_A$, $S'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}K_A$ and $S'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}K_A$ are given by

$$\begin{aligned}\sum_{q=1}^{T}\left(skit_{T-2}\right)_{tq}\left(k_A\right)_{tq} &= (t-1)\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho},\\ \sum_{q=1}^{T}\left(sikt_{T-2}\right)_{tq}\left(k_A\right)_{tq} &= \sum_{s=3}^{t}(s-2)\frac{1-\rho^{t-s+1}}{1-\rho}\end{aligned}$$

and

$$\sum_{q=1}^{T}\left(skkt_{T-2}\right)_{tq}\left(k_A\right)_{tq}=\left(t-1\right)\sum_{s=3}^{t}\left(s-2\right)\frac{1-\rho^{t-s+1}}{1-\rho},$$

respectively, for $3\leq t\leq T-1$. Therefore, by (D.6),

$$\begin{aligned}
V &= -\frac{S_2}{S_3}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}+\frac{S_1}{S_3}\sum_{t=3}^{T-1}\left(t-1\right)\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}\\
&\quad+\frac{S_1}{S_3}\sum_{t=3}^{T-1}\sum_{s=3}^{t}\left(s-2\right)\frac{1-\rho^{t-s+1}}{1-\rho}-\frac{S_0}{S_3}\sum_{t=3}^{T-1}\left(t-1\right)\sum_{s=3}^{t}\left(s-2\right)\frac{1-\rho^{t-s+1}}{1-\rho}\\
&= -A\sum_{t=3}^{T-1}\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}+B\sum_{t=3}^{T-1}t\sum_{s=3}^{t}\frac{1-\rho^{t-s+1}}{1-\rho}\\
&\quad+C\sum_{t=3}^{T-1}\sum_{s=3}^{t}s\frac{1-\rho^{t-s+1}}{1-\rho}-D\sum_{t=3}^{T-1}t\sum_{s=3}^{t}s\frac{1-\rho^{t-s+1}}{1-\rho}.
\end{aligned}$$

Similarly, by (D.7),

$$\begin{aligned}
W &= -\frac{S_2}{S_3}\mathrm{Tr}\left(S'_{T-2}\left(\iota_{T-2}\iota'_{T-2}\right)T_{T-2}L_A\right)+\frac{S_1}{S_3}\mathrm{Tr}\left(S'_{T-2}\left(\kappa_{T-2}\iota'_{T-2}\right)T_{T-2}L_A\right)\\
&\quad+\frac{S_1}{S_3}\mathrm{Tr}\left(S'_{T-2}\left(\iota_{T-2}\kappa'_{T-2}\right)T_{T-2}L_A\right)-\frac{S_0}{S_3}\mathrm{Tr}\left(S'_{T-2}\left(\kappa_{T-2}\kappa'_{T-2}\right)T_{T-2}L_A\right)\\
&= -A\sum_{t=3}^{T-1}\sum_{s=3}^{t}\rho^{t-s}+B\sum_{t=3}^{T-1}t\sum_{s=3}^{t}\rho^{t-s}+C\sum_{t=3}^{T-1}\sum_{s=3}^{t}s\rho^{t-s}-D\sum_{t=3}^{T-1}t\sum_{s=3}^{t}s\rho^{t-s}.
\end{aligned}$$

Application of the central limit theorem and Slutzky's theorem gives the desired result. ■

Remark. Analogously with (D.16), the probability limits P, Q, R, V and W for Models 2 and 3 can further be manipulated to obtain closed forms. These expressions are available upon request.

Bibliography

[1] Agnarsson, S., Carlin, P.S. (2002), "Family Background and the Estimated Return to Schooling", *Journal of Human Resources*, **37**(3), p.680-692.

[2] Ahn, H., Powell, J.L. (1993), "Semiparametric Estimation of Censored Selection Models", *Journal of Econometrics*, **58**(1-2), p.3-29.

[3] Ai, C., McFadden, D. (1997), "Estimation of Some Partially Specified Nonlinear Models", *Journal of Econometrics*, **76**(1-2), p.1-37.

[4] Akkina, K.R. (1974), "Application of Random Coefficient Regression Models to the Aggregation Problem", *Econometrica*, **42**, p.369-375.

[5] Amemiya, T., MaCurdy, T.E. (1986) "Instrumental-Variable Estimation of an Error-Components model", *Econometrica*, **54**(4), p.869-881.

[6] Anderson, T.W. (1970), "Estimation of Covariance Matrices Which are Linear Combinations or Whose Inverses are Linear Combinations of Given Matrices", in Bose, R.C. (ed.), *Essays in Probability and Statistics*, Chapel Hill, University of North Carolina Press, p.1-24.

[7] Andrews, D.W.K. (1991), "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", *Econometrica*, **59**(2), p.307-345.

[8] Angrist, J.D., Krueger, A.B. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, **106**(4), p.979-1014.

[9] Angrist, J.D., Newey, W.K. (1991), "Over-Identification Tests in Earnings Functions With Fixed Effects", *Journal of Business and Economic Statistics*, **9**(3), p.317-323.

[10] Ashenfelter, O., Rouse, C. (1998), "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins", *Quarterly Journal of Economics*, **113**(1), p.253-284.

[11] Baltagi, B.H., Kao, C.W. (2001), "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey", *Advances in Econometrics*, **15**, p.7-15.

[12] Baltagi, B. H., Khanti-Akom, S. (1990), "On Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators", *Journal of Applied Econometrics*, **5**(4), p.401-406.

[13] Banerjee, A. (1999), "Panel Data Unit Roots and Cointegration: An Overview", *Oxford Bulletin of Economics and Statistics*, Special Issue, p.607-629.

[14] Belsley, D.A., Kuh, E., Welsch, R.E. (1980), *Regression Diagnostics : Identifying Influential Data and Sources of Collinearity*, Wiley, N.Y.

[15] Belzil, C., Hansen, J. (2002),"A Structural Analysis of the Correlated Random Coefficient Wage Regression Model", *CIRANO Série Scientifique*, **2002s-07**, Centre interuniversitaire de recherche en analyse des organisations, Montréal, pp.33.

[16] Breusch, T.S., Mizon, G.E., Schmidt, P. (1989), "Efficient Estimation Using Panel Data", *Econometrica*, **57**(3), p.695-700.

[17] Card, D. (2001), "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems", *Econometrica*, **69**(5), p.1127-1160.

[18] Carroll, R.J., Ruppert, D. (1982), "Robust Estimation in Heteroskedastic Linear Models", *Annals of Statistics*, **10**(2), p.429-441.

[19] Chen, H. (1988), "Convergence Rates for Parametric Components in a Partially Linear Model", *Annals of Statistics*, **16**(1), p.136-146.

[20] Chen, S. (1997), "Semiparametric Estimation of the Type-3 Tobit Model", *Journal of Econometrics*, **80**(1), p.1-34.

[21] Choi, I. (2001), "Unit Root Tests for Panel Data", *Journal of International Money and Finance*, **20**, p.249-272.

[22] Christofides, L.N., Li, Q., Liu, Z.J. et al. (2003), "Recent Two-stage Sample Selection Procedures with an Application to the Gender Wage Gap", *Journal of Bussines and Economic Statistics*, **21**(3), p.396-405.

[23] Cornwell, C., Rupert, P. (1988), "Efficient Estimation with Panel Data: an Empirical Comparison of Instrumental Variables Estimators", *Journal of Applied Econometrics*, **3**, p.149-155.

[24] Cox, D.R., Snell, E.J. (1968), "A General Definition of Residuals", *Journal of the Royal Statistical Society B*, **30**(2), p.248-275.

[25] Crockett, P.W. (1985), "Asymptotic Distribution of the Hildreth-Houck Estimator", *Journal of the American Statistical Association*, **80**, p.202-204.

[26] Das, M. (2005), "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors", *Journal of Econometrics*, **124**(2), p.335-361.

[27] Das, M., Newey, W.K., Vella, F. (2003), "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies*, **70**(1), p.33-58.

[28] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

[29] Dickey, D. A., Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, **74**, p.427-431.

[30] Donald, S.G., Newey, W.K. (1994), "Series Estimation of Semilinear Regression", *Journal of Multivariate Analysis*, **50**(1), p.30-40.

[31] Fan, J., Härdle, W., Mammen, E. (1998), "Direct Estimation of Low Dimensional Components in Additive Models", *Annals of Statistics*, **26**(3), p.943-971.

[32] Fan, Y., Li, Q., Stengos, T. (1995), "Root-N-Consistent Semiparametric Regression with Conditional Heteroskedastic Disturbances", *Journal of Quantitative Economics*, **11**(1), p.229-240.

[33] Fisher, R.A. (1932), *Statistical Methods for Research Workers*, Oliver and Boyd, London.

[34] Frisch, R., Waugh, F.V. (1933), "Partial Time Regressions as Compared with Individual Trends", *Econometrica*, **1**(4), p.387-401.

[35] Fuller, W., Battese, G. (1973), "Transformations for Estimation of Linear Models with Error Structure", *Journal of the American Statistical Association*, **68**, p.626-632.

[36] Garen, J. (1984), "The Returns to Schooling: a Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica*, **52**(5), p.1199-1218.

[37] Gouriéroux, C., Monfort, A., Renault, E., Trognon, A. (1987), "Generalized Residuals", *Journal of Econometrics*, **34**(1-2), p.5-32.

[38] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

[39] Härdle, W., Liang, H., Gao, J. (2000), *Partially Linear Models*, Contributions to Statistics, Physica-Verlag, Heidelberg.

[40] Harmon, C., Hogan, V., Walker, I. (2003), "Dispersion in the Economic Return to Schooling", *Labour Economics*, **10**, p.205-214.

[41] Harmon, C., Walker, I. (1995), "Estimates of the Economic Return to Schooling for the United Kingdom", *American Economic Review*, **85**(5), p.1278-1286.

[42] Harris, R.D.F, Tzavalis, E. (1998), "Inference for Unit Roots in Dynamic Panels with Heteroscedastic and Serially Correlated Errors", *mimeo*, University of Exeter, Exeter.

[43] Harris, R.D.F, Tzavalis, E. (1999), "Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed", *Journal of Econometrics*, **91**, p.201-226.

[44] Hausman, J.A., Taylor, W.E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, **49**(6), p.1377-1398.

[45] Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement*, **5**(4), p.475-492.

[46] Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, **47**(1), p.153-161.

[47] Heckman, J., Tobias, J.L., Vytlacil, E. (2001), "Four Parameters of Interest in the Evaluation of Social Programs", *Southern Economic Journal*, **68**(2), p.210-224.

[48] Heckman, J., Vytlacil, E. (1998), "Instrumental Variables Methods for The Correlated Random Coefficient Model", *Journal of Human Resources*, **33**, p.974-987.

[49] Heckman, J., Vytlacil, E. (2001), "Identifying the Role of Cognitive Ability in Explaining the Level of and Change in the Return to Schooling", *Review of Economics and Statistics*, **83**(1), p.1-12.

[50] Hildreth, C.,Houck, J.P. (1968), "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, **63**, p.584-595.

[51] Hsiao, C. (1974), "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects", *International Economic Review*, **15**, p.12-30.

[52] Hsiao, C. (1975), "Some Estimation Methods for a Random Coefficient Model", *Econometrica*, **43**, p.305-325.

[53] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models", *Journal of Econometrics*, **58**(1-2), p.71-120.

[54] Im, K.S., Pesaran, M.H., Shin, Y. (2003), "Testing for Unit Roots in Heterogeneous Panels", *Journal of Econometrics*, **115**, p.53-74.

[55] Jennrich, R. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics*, **40**(2), p.633-643.

[56] Kane, T.J., Rouse, C.E. (1995), "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?", *American Economic Review*, **85**(3), p.600-614.

[57] Kelejian, H.H., Stephan, S.W. (1983), "Inference in Random Coefficient Panel Data Models: A Correction and Clarification of the Literature", *International Economic Review*, **24**(1), p.249-254.

[58] Kiviet, J. F. (1995), "On bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models", *Journal of Econometrics*, **68**, p.53-78.

[59] Kruiniger, H., Tzavalis, E. (2002),"Testing for Unit Roots in Short Dynamic Panels with Serially Correlated and Heteroscedastic Disturbance Terms", *mimeo*, Department of Economics, Queen Mary, University of London.

[60] Lauer, C. (2003), "Family Background, Cohort and Education: A French-German Comparison Based on a Multivariate Ordered Probit Model of Educational Attainment", *Labour Economics*, **10**, p.231-251.

[61] Lee, L.F. (1994), "Semiparametric Two-Stage Estimation of Sample Selection Models Subject to Tobit-Type Selection Variable", *Journal of Econometrics*, **61**(1), p.305-344.

[62] Lemke, R.J., Rischall, I.C. (2003), "Skill, Parental Income, and IV Estimation of the Returns to Schooling", *Applied Economics Letters*, **10**(5), p.281-286.

[63] Levin, A., Lin, C.-F. (1993), "Unit Root Tests in Panel Data: New Results", *mimeo*, University of California, San Diego.

[64] Levin, A., Lin, C.-F., Chu, C.-S. (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties" *Journal of Econometrics*, **108**(1), p.1-24.

[65] Li, Q. (1996), "On the Root-N-consistent Semiparametric Estimation of Partially Linear Models", *Economics Letters*, **51**(3), p.277-285.

[66] Linton, O. (1997), "Efficient Estimation of Additive Nonparametric Regression Models", *Biometrika*, **84**(2), p.469-473.

[67] Lovell, M.C. (1963), "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis", *Journal of the American Statistical Association*, **58**(304), p.993-1010.

[68] Maddala, G.S., Wu, S. (1999), "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test", *Oxford Bulletin of Economics and Statistics*, Special Issue, p.631-652.

[69] Mandy, D.M., Martins-Filho, C. (1994), "A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variables Estimators", *International Economic Review*, **35**(4), p.957-979.

[70] Martins, P.S., Pereira, P.T. (2004), "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries", *Labour Economics*, **11**(3), p.355-371.

[71] Mincer, J. (1974), *Schooling, Experience and Earnings*, Columbia University Press, New York.

[72] Murphy, K.M., Topel, R.H. (1985), "Estimation and Inference in Two-Step Econometric Models", *Journal of Business and Economic Statistics*, **3**(4), p.370-379.

[73] Newey, W.K. (1997), "Convergence Rates and Asymptotic Normality of Series Estimators", *Journal of Econometrics*, **79**(1), p.147-168.

[74] Newey, W. (1999), "Two-Step Series Estimation of Sample Selection Models", *mimeo*, MIT.

[75] Newey, W., Powell, J., Vella, F. (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, **67**(3), p.565-603.

[76] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, **49**, p.1417-1426.

[77] Pesaran, M.H., Smith, R. (1995), "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels", *Journal of Econometrics*, **68**, p.79-113.

[78] Phillips, P.C.B., Moon, H.R. (1999a), "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, **67**, p.1057-1111.

[79] Phillips, P.C.B., Moon, H.R. (1999b), "Nonstationary Panel Data Analysis: An Overview of some Recent Developments ", *mimeo*, Cowles Foundation for Research in Economics, Yale University.

[80] Phillips, P.C.B., Perron, P. (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, **75**(2), p.335-346.

[81] Pierce, D.A., 1982, The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics, *Annals of Statistics* **10**(2), 475-478.

[82] Powell, J.L. (1994), "Estimation of Semiparametric Models", in Engle, R.F., McFadden, D. (eds.), *Handbook of Econometrics*, **Vol. 4**, Elsevier, Amsterdam.

[83] Quah, D. (1994), "Exploiting Cross-section Variations for Unit Root Inference in Dynamic Data", *Economics Letters*, **44**, p.9-19.

[84] Randles, R.H., 1982, On the Asymptotic Normality of Statistics with Estimated Parameters, *Annals of Statistics* **10**(2), 462-474.

[85] Rao, C.R. (1965), "The Theory of Least Squares When the Parameters Are Stochastic and Its Application to the Analysis of Growth Curves", *Biometrika*, **52**, p.447-458.

[86] *R* Development Core Team (2004), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[87] Resende, M. (2004), "Gibrat's Law and the Growth of Cities in Brazil: A Panel Data Investigation", *Urban Studies*, **41**(8), p.1537-1549.

[88] Robinson, P. (1988), "Root-N Consistent Semiparametric Regression", *Econometrica*, **56**(4), p.931-954.

[89] Röhrig, C.S. (1988), "Conditions for Identification in Nonparametric and Parametric Models", *Econometrica*, **56**(2), p.433-447.

[90] Rosenblatt, M. (1956), "Remarks on Some Non-Parametric Estimates of a Density Function", *Annals of Mathematical Statistics*, **27**, p.642-669.

[91] Searle, S.R., Casella, G., McCulloch, C.E. (1992), *Variance Components*, Wiley Series in Probability and Mathematical Statistics, Applied Probability and Statistics, Wiley, N.Y.

[92] Sen, A.K., Srivastava, M. (1990), *Regression Analysis: Theory, Methods and Applications*, Springer Texts in Statistics, Springer-Verlag, New York.

[93] Stone, C.J. (1985), "Additive Regression and Other Nonparametric Models", *Annals of Statistics*, **13**(2), p.689-705.

[94] Swamy, P.A.V.B. (1970), "Efficient Inference in a Random Coefficient Regression Model", *Econometrica*, **38**(2), p.311-323.

[95] Szegö, G. (1967), *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, **23**.

[96] Tobias, J.L. (2003), "Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability-earnings Relationships", *Oxford Bulletin of Economics and Statistics*, **65**(1), p.1-29.

[97] Vella, F. (1992), "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models", *Journal of Applied Econometrics*, **7**(4), p.413-421.

[98] Vella, F. (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors", *International Economic Review*, **34**(2), p.441-457.

[99] Vella, F. (1998), "Estimating Models with Sample Selection Bias: A Survey", *Journal of Human Resources*, **33**(1), p.127-169.

[100] Vella, F., Verbeek, M. (1999), "Two-Step Estimation of Simultaneous Equation Panel Data Models with Censored Endogenous Variables", *Journal of Econometrics*, **90**(2), p.239-263.

[101] Wald, A. (1947), "A Note on Regression Analysis", *Annals of Mathematical Statistics*, **18**, p.586-589.

[102] White, H. (1980a), "Using Least Squares to Approximate Unknown Regression Functions", *International Economic Review*, **21**(1), p.149-170.

[103] White, H. (1980b), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.", *Econometrica*, **48**(4), p.817-838.

[104] White, H. (1984), *Asymptotic Theory for Econometricians*, Orlando, Academic Press.

[105] Wooldridge, J.M. (1994), "Selection Corrections With a Censored Selection Variable", *mimeo*, Michigan State University.

[106] Wooldridge, J.M. (1997), "On two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model", *Economics Letters*, **56**, p.129-133.

[107] Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Mass.

[108] Wooldridge, J.M. (2003a), "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model", *Economics Letters*, **79**, p.129-133.

[109] Wooldridge, J.M. (2003b), "On the Robustness of Fixed Effects and Related Estimators in Correlated Random Coefficient Panel Data Models", *mimeo*.

[110] Wyhowski, D.J. (1994), "Estimation of a Panel-Data Model in the Presence of Correlation between Regressors and a 2-Way Error Component", *Econometric Theory*, **10**, p. 130-139.

[111] Yatchew, A. (2003), *Semiparametric Regression for the Applied Econometrician*, Themes in Modern Econometrics, Cambridge University Press.

[112] Zellner, A. (1969), "On the Aggregation Problem: A New Approach to a Troublesome Problem", in Fox, K.A. et al. (eds.) *Economic Models, Estimation and Risk Programming: Essays in Honor of Gerhard Tintner*, Springer-Verlag, p.365-378.

List of Doctoral Dissertations

Doctoral dissertations from the Faculty of Economic and Applied Economic Sciences (from August 1, 1971)

1. Gepts, Stefaan *Stability and efficiency of resource allocation processes in discrete commodity spaces.* Leuven, KUL, 1971. 86 pp.
2. Peeters, Theo *Determinanten van de internationale handel in fabrikaten.* Leuven, Acco, 1971. 290 pp.
3. Van Looy, Wim *Personeelsopleiding: een onderzoek naar investeringsaspekten van opleiding.* Hasselt, Vereniging voor wetenschappelijk onderzoek in Limburg, 1971. VII, 238 pp.
4. Tharakan, Mathew *Indian exports to the European community: problems and prospects.* Leuven, Faculty of economics and applied economics, 1972. X, 343 pp.
5. Herroelen, Willy *Heuristische programmatie: methodologische benadering en praktische toepassing op complexe combinatorische problemen.* Leuven, Aurelia scientifica, 1972. X, 367 pp.
6. Vandenbulcke, Jacques *De studie en de evaluatie van data-organisatiemethodes en datazoekmethodes.* Leuven, s.n., 1973. 3 V.
7. Pennycuick, Roy A. *The economics of the ecological syndrome.* Leuven, Acco, 1973. XII, 177 pp.
8. Kawata, T. Bualum *Formation du capital d'origine belge, dette publique et stratégie du développement au Zaire.* Leuven, KUL, 1973. V, 342 pp.
9. Donckels, Rik *Doelmatige oriëntering van de sectorale subsidiepolitiek in België: een theoretisch onderzoek met empirische toetsing.* Leuven, K.U.Leuven, 1974. VII, 156 pp.
10. Verhelst, Maurice *Contribution to the analysis of organizational information systems and their financial benefits.* Leuven, K.U.Leuven, 1974. 2 V.
11. Clemeur, Hugo *Enkele verzekeringstechnische vraagstukken in het licht van de nutstheorie.* Leuven, Aurelia scientifica, 1974. 193 pp.

12. Heyvaert, Edward *De ontwikkeling van de moderne bank- en krediettechniek tijdens de zestiende en zeventiende eeuw in Europa en te Amsterdam in het bijzonder.* Leuven, K.U.Leuven, 1975. 186 pp.

13. Vertonghen, Robert *Investeringscriteria voor publieke investeringen: het uitwerken van een operationele theorie met een toepassing op de verkeersinfrastructuur.* Leuven, Acco, 1975. 254 pp.

14. Niet toegekend.

15. Vanoverbeke, Lieven *Microeconomisch onderzoek van de sectoriële arbeidsmobiliteit.* Leuven, Acco, 1975. 205 pp.

16. Daems, Herman *The holding company: essays on financial intermediation, concentration and capital market imperfections in the Belgian economy.* Leuven, K.U.Leuven, 1975. XII, 268 pp.

17. Van Rompuy, Eric *Groot-Brittannië en de Europese monetaire integratie: een onderzoek naar de gevolgen van de Britse toetreding op de geplande Europese monetaire unie.* Leuven, Acco, 1975. XIII, 222 pp

18. Moesen, Wim *Het beheer van de staatsschuld en de termijnstructuur van de intrestvoeten met een toepassing voor België.* Leuven, Vander, 1975. XVI, 250 pp.

19. Lambrecht, Marc *Capacity constrained multi-facility dynamic lot-size problem.* Leuven, KUL, 1976. 165 pp.

20. Raymaeckers, Erik *De mens in de onderneming en de theorie van het producentengedrag: een bijdrage tot transdisciplinaire analyse.* Leuven, Acco, 1976. XIII, 538 pp.

21. Tejano, Albert *Econometric and input-output models in development planning: the case of the Philippines.* Leuven, KUL, 1976. XX, 297 pp.

22. Martens, Bernard *Prijsbeleid en inflatie met een toepassing op België.* Leuven, KUL, 1977. IV, 253 pp.

23. Verheirstraeten, Albert *Geld, krediet en intrest in de Belgische financiële sector.* Leuven, Acco, 1977. XXII, 377 pp.

24. Gheyssens, Lieven *International diversification through the government bond market: a risk-return analysis.* Leuven, s.n., 1977. 188 pp.

25. Lefebvre, Chris *Boekhoudkundige verwerking en financiële verslaggeving van huurkooptransacties en verkopen op afbetaling bij ondernemingen die aan consumenten verkopen.* Leuven, KUL, 1977. 228 pp.

26. Kesenne, Stefan *Tijdsallocatie en vrijetijdsbesteding: een econometrisch onderzoek.* Leuven, s.n., 1978. 163 pp.

27. Van Herck, Gustaaf *Aspecten van optimaal bedrijfsbeleid volgens het marktwaardecriterium: een risico-rendementsanalyse.* Leuven, KUL, 1978. IV, 163 pp.

28. Van Poeck, Andre *World price trends and price and wage development in Belgium: an investigation into the relevance of the Scandinavian model of inflation for Belgium.* Leuven, s.n., 1979. XIV, 260 pp.

29. Vos, Herman *De industriële technologieverwerving in Brazilië: een analyse.* Leuven, s.n., 1978. onregelmatig gepagineerd.

30. Dombrecht, Michel *Financial markets, employment and prices in open economies.* Leuven, KUL, 1979. 182 pp.

31. De Pril, Nelson *Bijdrage tot de actuariële studie van het bonus-malussysteem.* Brussel, OAB, 1979. 112 pp.

32. Carrin, Guy *Economic aspects of social security: a public economics approach.* Leuven, KUL, 1979. onregelmatig gepagineerd

33. Regidor, Baldomero *An empirical investigation of the distribution of stock-market prices and weak-form efficiency of the Brussels stock exchange.* Leuven, KUL, 1979. 214 pp.

34. De Groot, Roger *Ongelijkheden voor stop loss premies gebaseerd op E.T. systemen in het kader van de veralgemeende convexe analyse.* Leuven, KUL, 1979. 155 pp.

35. Ceyssens, Martin *On the peak load problem in the presence of rationizing by waiting.* Leuven, KUL, 1979. IX, 217 pp.

36. Abdul Razk Abdul *Mixed enterprise in Malaysia: the case study of joint venture between Malysian public corporations and foreign enterprises.* Leuven, KUL, 1979. 324 pp.

37. De Bruyne, Guido *Coordination of economic policy: a game-theoretic approach.* Leuven, KUL, 1980. 106 pp.

38. Kelles, Gerard *Demand, supply, price change and trading volume on financial markets of the matching-order type.* (Vraag, aanbod, koersontwikkeling en omzet op financiële markten van het Europese type.) Leuven, KUL, 1980. 222 pp.

39. Van Eeckhoudt, Marc *De invloed van de looptijd, de coupon en de verwachte inflatie op het opbrengstverloop van vastrentende finaciële activa.* Leuven, KUL, 1980. 294 pp.

40. Sercu, Piet *Mean-variance asset pricing with deviations from purchasing power parity.* Leuven, s.n., 1981. XIV, 273 pp.

41. Dequae, Marie-Gemma *Inflatie, belastingsysteem en waarde van de onderneming.* Leuven, KUL, 1981. 436 pp.

42. Brennan, John *An empirical investigation of Belgian price regulation by prior notification: 1975 – 1979 – 1982.* Leuven, KUL, 1982. XIII, 386 pp.

43. Colla, Annie *Een econometrische analyse van ziekenhuiszorgen.* Leuven, KUL, 1982. 319 pp.

44. Niet toegekend.

45. Schokkaert, Eric *Modelling consumer preference formation.* Leuven, KUL, 1982. VIII, 287 pp.

46. Degadt, Jan *Specificatie van een econometrisch model voor vervuilingsproblemen met proeven van toepassing op de waterverontreiniging in België.* Leuven, s.n., 1982. 2 V.

47. Lanjong, Mohammad Nasir *A study of market efficiency and risk-return relationships in the Malaysian capital market.* s.l., s.n., 1983. XVI, 287 pp.

48. Proost, Stef *De allocatie van lokale publieke goederen in een economie met een centrale overheid en lokale overheden.* Leuven, s.n., 1983. onregelmatig gepagineerd.

49. Van Hulle, Cynthia (/08/83) *Shareholders' unanimity and optimal corporate decision making in imperfect capital markets.* s.l., s.n., 1983. 147 pp. + appendix.

50. Van Wouwe, Martine (2/12/83) *Ordening van risico's met toepassing op de berekening van ultieme ruïnekansen.* Leuven, s.n., 1983. 109 pp.

51. D'Alcantara, Gonzague (15/12/83) *SERENA: a macroeconomic sectoral regional and national account econometric model for the Belgian economy.* Leuven, KUL, 1983. 595 pp.

52. D'Have, Piet (24/02/84) *De vraag naar geld in België.* Leuven, KUL, 1984. XI, 318 pp.

53. Maes, Ivo (16/03/84) *The contribution of J.R. Hicks to macro-economic and monetary theory.* Leuven, KUL, 1984. V, 224 pp.

54. Subianto, Bambang (13/09/84) *A study of the effects of specific taxes and subsidies on a firms' R&D investment plan.* s.l., s.n., 1984. V, 284 pp.

55. Sleuwaegen, Leo (26/10/84) *Location and investment decisions by multinational enterprises in Belgium and Europe.* Leuven, KUL, 1984. XII, 247 pp.

56. Geyskens, Erik (27/03/85) *Produktietheorie en dualiteit.* Leuven, s.n., 1985. VII, 392 pp.

57. Cole, Frank (26/06/85) *Some algorithms for geometric programming.* Leuven, KUL, 1985. 166 pp.

58. Standaert, Stan (26/09/86) *A study in the economics of repressed consumption.* Leuven, KUL, 1986. X, 380 pp.

59. Delbeke, Jos (03/11/86) *Trendperioden in de geldhoeveelheid van België 1877-1983: een theoretische en empirische analyse van de "Banking school" hypothese.* Leuven, KUL, 1986. XII, 430 pp.

60. Vanthienen, Jan (08/12/86) *Automatiseringsaspecten van de specificatie, constructie en manipulatie van beslissingstabellen.* Leuven, s.n., 1986. XIV, 378 pp.

61. Luyten, Robert (30/04/87) *A systems-based approach for multi-echelon production/ inventory systems.* s.l., s.n., 1987. 3V.

62. Mercken, Roger (27/04/87) *De invloed van de data base benadering op de interne controle.* Leuven, s.n., 1987. XIII, 346 pp.

63. Van Cayseele, Patrick (20/05/87) *Regulation and international innovative activities in the pharmaceutical industry.* s.l., s.n., 1987. XI, 169 pp.

64. François, Pierre (21/09/87) *De empirische relevantie van de 'independence from irrelevant alternatives' assumptie in discrete keuzemodellen.* Leuven, s.n., 1987. IX, 379 pp.

65. Decoster, André (23/09/88) *Family size, welfare and public policy.* Leuven, KUL. Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. XIII, 444 pp.

66. Heijnen, Bart (09/09/88) *Risicowijziging onder invloed van vrijstellingen en herverzekeringen: een theoretische analyse van optimaliteit en premiebepaling.* Leuven, KUL. Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. onregelmatig gepagineerd.

67. Geeroms, Hans (14/10/88) *Belastingvermijding. Theoretische analyse van de determinanten van de belastingontduiking en de belastingontwijking met empirische verificaties.* Leuven, s.n., 1988. XIII, 409, 5 pp.

68. Put, Ferdi (19/12/88) *Introducing dynamic and temporal aspects in a conceptual (database) schema.* Leuven, KUL. Faculteit Economische en Toegepaste Economische Wetenschappen, 1988. XVIII, 415 pp.

69. Van Rompuy, Guido (13/01/89) *A supply-side approach to tax reform programs. Theory and empirical evidence for Belgium.* Leuven, KUL. Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XVI, 189, 6 pp.

70. Peeters, Ludo (19/06/89) *Een ruimtelijk evenwichtsmodel van de graanmarkten in de E.G.: empirische specificatie en beleidstoepassingen.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XVI, 412 pp.

71. Pacolet, Jozef (10/11/89) *Marktstructuur en operationele efficiëntie in de Belgische financiële sector.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. XXII, 547 pp.

72. Vandenbroek, Martina (13/12/89) *Optimalisatie van verzekeringscontracten en premieberekeningsprincipes.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1989. 95 pp.

73. Willekens, François () *Determinants of government growth in industrialized countries with applications to Belgium.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. VI, 332 pp.

74. Veugelers, Reinhilde (02/04/90) *Scope decisions of multinational enterprises.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. V, 221 pp.

75. Kesteloot, Katrien (18/06/90) *Essays on performance diagnosis and tacit cooperation in international oligopolies.* Leuven, K.U.Leuven. Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. 227 pp.

76. Wu, Changqi (23/10/90) *Strategic aspects of oligopolistic vertical integration.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1990. VIII, 222 pp.

77. Zhang, Zhaoyong (08/07/91) *A disequilibrium model of China's foreign trade behaviour.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1991. XII, 256 pp.

78. Dhaene, Jan (25/11/91) *Verdelingsfuncties, benaderingen en foutengrenzen van stochastische grootheden geassocieerd aan verzekeringspolissen en -portefeuilles.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1991. 146 pp.

79. Bauwelinckx, Thierry (07/01/92) *Hierarchical credibility techniques.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 130 pp.

80. Demeulemeester, Erik (23/3/92) *Optimal algorithms for various classes of multiple resource-constrained project scheduling problems.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 180 pp.

81. Steenackers, Anna (1/10/92) *Risk analysis with the classical actuarial risk model: theoretical extensions and applications to Reinsurance.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 139 pp.

82. Cockx, Bart (24/09/92) *The minimum income guarantee. Some views from a dynamic perspective.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. XVII, 401 pp.

83. Meyermans, Eric (06/11/92) *Econometric allocation systems for the foreign exchange market: Specification, estimation and testing of transmission mechanisms under currency substitution.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. XVIII, 343 pp.

84. Chen, Guoqing (04/12/92) *Design of fuzzy relational databases based on fuzzy functional dependency.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1992. 176 pp.

85. Claeys, Christel (18/02/93) *Vertical and horizontal category structures in consumer decision making: The nature of product hierarchies and the effect of brand typicality.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 348 pp.

86. Chen, Shaoxiang (25/03/93) *The optimal monitoring policies for some stochastic and dynamic production processes.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 170 pp.

87. Overweg, Dirk (23/04/93) *Approximate parametric analysis and study of cost capacity management of computer configurations.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 270 pp.

88. Dewachter, Hans (22/06/93) *Nonlinearities in speculative prices: The existence and persistence of nonlinearity in foreign exchange rates.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 151 pp.

89. Lin, Liangqi (05/07/93) *Economic determinants of voluntary accounting choices for R & D expenditures in Belgium.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 192 pp.

90. Dhaene, Geert (09/07/93) *Encompassing: formulation, properties and testing.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 117 pp.

91. Lagae, Wim (20/09/93) *Marktconforme verlichting van soevereine buitenlandse schuld door private crediteuren: een neo-institutionele analyse.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 241 pp.

92. Van De Gaer, Dirk (27/09/93) *Equality of opportunity and investment in human capital.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1993. 172 pp.

93. Schroyen, Alfred (28/02/94) *Essays on redistributive taxation when monitoring is costly.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 203 pp. + V.

94. Steurs, Geert (15/07/94) *Spillovers and cooperation in research and development.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 266 pp.

95. Baras, Johan (15/09/94) *The small sample distribution of the Wald, Lagrange multiplier and likelihood ratio tests for homogeneity and symmetry in demand analysis: a Monte Carlo study.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 169 pp.

96. Gaeremynck, Ann (08/09/94) *The use of depreciation in accounting as a signalling device.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 232 pp.

97. Bettendorf, Leon (22/09/94) *A dynamic applied general equilibrium model for a small open economy.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 149 pp.

98. Teunen, Marleen (10/11/94) *Evaluation of interest randomness in actuarial quantities.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1994. 214 pp.

99. Van Ootegem, Luc (17/01/95) *An economic theory of private donations.* Leuven. K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 236 pp.

100. De Schepper, Ann (20/03/95) *Stochastic interest rates and the probabilistic behaviour of actuarial functions.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 211 pp.

101. Lauwers, Luc (13/06/95) *Social choice with infinite populations.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 79 pp.

102. Wu, Guang (27/06/95) *A systematic approach to object-oriented business modeling.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 248 pp.

103. Wu, Xueping (21/08/95) *Term structures in the Belgian market: model estimation and pricing error analysis.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 133 pp.

104. Pepermans, Guido (30/08/95) *Four essays on retirement from the labor force.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 128 pp.

105. Algoed, Koen (11/09/95) *Essays on insurance: a view from a dynamic perspective.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 136 pp.

106. Degryse, Hans (10/10/95) *Essays on financial intermediation, product differentiation, and market structure.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 218 pp.

107. Meir, Jos (05/12/95) *Het strategisch groepsconcept toegepast op de Belgische financiële sector.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1995. 257 pp.

108. Wijaya, Miryam Lilian (08/01/96) *Voluntary reciprocity as an informal social insurance mechanism: a game theoretic approach.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 124 pp.

109. Vandaele, Nico (12/02/96) *The impact of lot sizing on queueing delays: multi product, multi machine models.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 243 pp.

110. Gielens, Geert (27/02/96) *Some essays on discrete time target zones and their tails.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 131 pp.

111. Guillaume, Dominique (20/03/96) *Chaos, randomness and order in the foreign exchange markets. Essays on the modelling of the markets.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 171 pp.

112. Dewit, Gerda (03/06/96) *Essays on export insurance subsidization.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 186 pp.

113. Van Den Acker, Carine (08/07/96) *Belief-function theory and its application to the modeling of uncertainty in financial statement auditing.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 147 pp.

114. Imam, Mahmood Osman (31/07/96) *Choice of IPO Flotation Methods in Belgium in an Asymmetric Information Framework and Pricing of IPOs in the Long-Run.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 221 pp.

115. Nicaise, Ides (06/09/96) *Poverty and Human Capital.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1996. 209 pp.

116. Eyckmans, Johan (18/09/97) *On the Incentives of Nations to Join International Environmental Agreements.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1997. XV + 348 pp.

117. Crisologo-Mendoza, Lorelei (16/10/97) *Essays on Decision Making in Rural Households: a study of three villages in the Cordillera Region of the Philippines.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1997. 256 pp.

118. De Reyck, Bert (26/01/98) *Scheduling Projects with Generalized Precedence Relations: Exact and Heuristic Procedures.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. XXIV + 337 pp.

119. Vandemaele, Sigrid (30/04/98) *Determinants of Issue Procedure Choice within the Context of the French IPO Market: Analysis within an Asymmetric Information Framework.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 241 pp.

120. Vergauwen, Filip (30/04/98) *Firm Efficiency and Compensation Schemes for the Management of Innovative Activities and Knowledge Transfers.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. VIII + 175 pp.

121. Leemans, Herlinde (29/05/98) *The Two-Class Two-Server Queueing Model with Non-preemptive Heterogeneous Priority Structures.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 211 pp.

122. Geyskens, Inge (4/09/98) *Trust, Satisfaction, and Equity in Marketing Channel Relationships.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 202 pp.

123. Sweeney, John (19/10/98) *Why Hold a Job ? The Labour Market Choice of the Low-Skilled.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1998. 278 pp.

124. Goedhuys, Micheline (17/03/99) *Industrial Organisation in Developing Countries, Evidence from Côte d'Ivoire.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 251 pp.

125. Poels, Geert (16/04/99) *On the Formal Aspects of the Measurement of Object-Oriented Software Specifications.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 507 pp.

126. Mayeres, Inge (25/05/99) *The Control of Transport Externalities: A General Equilibrium Analysis.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. XIV + 294 pp.

127. Lemahieu, Wilfried (5/07/99) *Improved Navigation and Maintenance through an Object-Oriented Approach to Hypermedia Modelling.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 284 pp.

128. Van Puyenbroeck, Tom (8/07/99) *Informational Aspects of Fiscal Federalism.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 192 pp.

129. Van Den Poel, Dirk (5/08/99) *Response Modeling for Database Marketing Using Binary Classification.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 342 pp.

130. Gielens, Katrijn (27/08/99) *International Entry Decisions in the Retailing Industry: Antecedents and Performance Consequences.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 336 pp.

131. Peeters, Anneleen (16/12/99) *Labour Turnover Costs, Employment and Temporary Work.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 207 pp.

132. Vanhoenacker, Jurgen (17/12/99) *Formalizing a Knowledge Management Architecture Meta-Model for Integrated Business Process Management.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 1999. 252 pp.

133. Nunes, Paulo (20/03/2000) *Contingent Valuation of the Benefits of Natural Areas and its Warmglow Component.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. XXI + 282 pp.

134. Van Den Cruyce, Bart (7/04/2000) *Statistische discriminatie van allochtonen op jobmarkten met rigide lonen.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. XXIII + 441 pp.

135. Repkine, Alexandre (15/03/2000) *Industrial restructuring in countries of Central and Eastern Europe: Combining branch-, firm- and product-level data for a better understanding of Enterprises' behaviour during transition towards market economy.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VI + 147 pp.

136. Aksoy, Yunus (21/06/2000) *Essays on international price rigidities and exchange rates.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. IX + 236 pp.

137. Riyanto, Yohanes Eko (22/06/2000) *Essays on the internal and external delegation of authority in firms.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VIII + 280 pp.

138. Huyghebaert, Nancy (20/12/2000) *The Capital Structure of Business Start-ups.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2000. VIII + 332 pp.

139. Franckx, Laurent (22/01/2001) *Ambient Inspections and Commitment in Environmental Enforcement.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001 VIII + 286 pp.

140. Vandille, Guy (16/02/2001) *Essays on the Impact of Income Redistribution on Trade.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001 VIII + 176 pp.

141. Marquering, Wessel (27/04/2001) *Modeling and Forecasting Stock Market Returns and Volatility.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. V + 267 pp.

142. Faggio, Giulia (07/05/2001) *Labor Market Adjustment and Enterprise Behavior in Transition.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 150 pp.

143. Goos, Peter (30/05/2001) *The Optimal Design of Blocked and Split-plot experiments.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001.X + 224 pp.

144. Labro, Eva (01/06/2001) *Total Cost of Ownership Supplier Selection based on Activity Based Costing and Mathematical Programming.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 217 pp.

145. Vanhoucke, Mario (07/06/2001) *Exact Algorithms for various Types of Project Scheduling Problems. Nonregular Objectives and time/cost Trade-offs.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 316 pp.

146. Bilsen, Valentijn (28/08/2001) *Entrepreneurship and Private Sector Development in Central European Transition Countries.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. XVI + 188 pp.

147. Nijs, Vincent (10/08/2001) *Essays on the dynamic Category-level Impact of Price promotions.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001.

148. Cherchye, Laurens (24/09/2001) *Topics in Non-parametric Production and Efficiency Analysis.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VII + 169 pp.

149. Van Dender, Kurt (15/10/2001) *Aspects of Congestion Pricing for Urban Transport.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. VII + 203 pp.

150. Capéau, Bart (26/10/2001) *In defence of the excess demand approach to poor peasants' economic behaviour. Theory and Empirics of non-recursive agricultural household modelling.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. XIII + 286 blz.

151. Calthrop, Edward (09/11/2001) *Essays in urban transport economics.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001.

152. Vander Bauwheide, Heidi (03/12/2001) *Earnings management in an Non-Anglo-Saxon environment.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2001. 408 pp.

153. De Backer, Koenraad (22/01/2002) *Multinational firms and industry dynamics in host countries : the case of Belgium.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VII + 165 pp.

154. Bouwen, Jan (08/02/2002) *Transactive memory in operational workgroups. Concept elaboration and case study.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 319 pp. + appendix 102 pp.

155. Van Den Brande, Inge (13/03/2002) *The psychological contract between employer and employee : a survey among Flemish employees.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VIII + 470 pp.

156. Veestraeten, Dirk (19/04/2002) *Asset Price Dynamics under Announced Policy Switching.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 176 pp.

157. Peeters, Marc (16/05/2002) *One Dimensional Cutting and Packing : New Problems and Algorithms.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

158. Skudelny, Frauke (21/05/2002) *Essays on The Economic Consequences of the European Monetary Union.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

159. De Weerdt, Joachim (07/06/2002) *Social Networks, Transfers and Insurance in Developing countries.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. VI + 129 pp.

160. Tack, Lieven (25/06/2002) *Optimal Run Orders in Design of Experiments.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XXXI + 344 pp.

161. Poelmans, Stephan (10/07/2002) *Making Workflow Systems work. An investigation into the Importance of Task-appropriation fit, End-user Support and other Technological Characteristics.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 237 pp.

162. Jans, Raf (26/09/2002) *Capacitated Lot Sizing Problems : New Applications, Formulations and Algorithms.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

163. Viaene, Stijn (25/10/2002) *Learning to Detect Fraud from enriched Insurance Claims Data (Context, Theory and Applications).* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 315 pp.

164. Ayalew, Tekabe (08/11/2002) *Inequality and Capital Investment in a Subsistence Economy.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. V + 148 pp.

165. Mues, Christophe (12/11/2002) *On the Use of Decision Tables and Diagrams in Knowledge Modeling and Verification.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. 222 pp.

166. Brock, Ellen (13/03/2003) *The Impact of International Trade on European Labour Markets.* K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002.

167. Vermeulen, Frederic (29/11/2002) *Essays on the collective Approach to Household Labour Supply.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XIV + 203 pp.

168. Cludts, Stephan (11/12/2002) *Combining participation in decision-making with financial participation : theoretical and empirical perspectives.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2002. XIV + 247 pp.

169. Warzynski, Frederic (09/01/2003) *The dynamic effect of competition on price cost margins and innovation.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen 2003.

170. Verwimp, Philip (14/01/2003) *Development and genocide in Rwanda ; a political economy analysis of peasants and power under the Habyarimana regime.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

171. Bigano, Andrea (25/02/2003) *Environmental regulation of the electricity sector in a European Market Framework.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. XX + 310 pp.

172. Maes, Konstantijn (24/03/2003) *Modeling the Term Structure of Interest Rates Across Countries.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. V+246 pp.

173. Vinaimont, Tom (26/02/2003) *The performance of One- versus Two-Factor Models of the Term Structure of Interest Rates.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

174. Ooghe, Erwin (15/04/2003) *Essays in multi-dimensional social choice.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. VIII+108 pp.

175. Forrier, Anneleen (25/04/2003) *Temporary employment, employability and training.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

176. Cardinaels, Eddy (28/04/2003) *The role of cost system accuracy in managerial decision making.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. 144 pp.

177. De Goeij, Peter (02/07/2003) *Modeling Time-Varying Volatility and Interest Rates.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003. VII+225 pp.

178. Leus, Roel (19/09/2003) *The generation of stable project plans. Complexity and exact algorithms.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

179. Marinheiro, Carlos (23/09/2003) *EMU and fiscal stabilisation policy : the case of small countries.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen 2003.

180. Baessens, Bart (24/09/2003) *Developing intelligent systems for credit scoring using machine learning techniques.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

181. Kóczy, László (18/09/2003) *Solution concepts and outsider behaviour in coalition formation games.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

182. Altomonte, Carlo (25/09/2003) *Essays on Foreign Direct Investment in transition countries : learning from the evidence.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

183. Dries, Liesbeth (10/11/2003) *Transition, Globalisation and Sectoral Restructuring: Theory and Evidence from the Polish Agri-Food Sector.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

184. Devooght, Kurt (18/11/2003) *Essays On Responsibility-Sensitive Egalitarianism and the Measurement of Income Inequality.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

185. Deleersnyder, Barbara (28/11/2003) *Marketing in Turbulent Times.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

186. Ali, Daniel (19/12/2003) *Essays on Household Consumption and Production Decisions under Uncertainty in Rural Ethiopia.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2003.

187. Willems, Bert (14/01/2004) *Electricity networks and generation market power.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

188. Janssens, Gust (30/01/2004) *Advanced Modelling of Conditional Volatility and Correlation in Financial Markets.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

189. Thoen, Vincent (19/01/2004) *On the valuation and disclosure practices implemented by venture capital providers.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

190. Martens, Jurgen (16/02/2004) *A fuzzy set and stochastic system theoretic technique to validate simulation models.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

191. Altavilla, Carlo (21/05/2004) *Monetary policy implementation and transmission mechanisms in the Euro area.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

192. De Bruyne, Karolien (07/06/2004) *Essays in the location of economic activity.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

193. Adem, Jan (25/06/2004) *Mathematical programming approaches for the supervised classification problem.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

194. Lerouge, Davy (08/07/2004) *Predicting Product Preferences : the effect of internal and external cues.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

195. Vandenbroeck, Katleen (16/07/2004) *Essays on output growth, social learning and land allocation in agriculture : micro-evidence from Ethiopia and Tanzania.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

196. Grimaldi, Maria (03/09/004) *The exchange rate, heterogeneity of agents and bounded rationality.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

197. Smedts, Kristien (26/10/2004) *Financial integration in EMU in the framework of the no-arbitrage theory.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

198. Koevoets, Wim (12/11/2004) *Essays on Unions, Wages and Employment.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

199. Callens, Marc (22/11/2004) *Essays on multilevel logistic Regression.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

200. Ruggoo, Arvind (13/12/2004) *Two stage designs robust to model uncertainty.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2004.

201. Hoorelbeke, Dirk (28/01/2005) *Bootstrap and Pivoting Techniques for Testing Multiple Hypotheses.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

202. Rousseau, Sandra (17/02/2005) *Selecting Environmental Policy Instruments in the Presence of Incomplete Compiance.*, Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

203. Van Der Meulen, Sofie (17/02/2005) *Quality of Financial Statements : Impact of the external auditor and applied accounting standards.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

204. Dimova, Ralitza (21/02/2005) *Winners and Losers during Structural Reform and Crisis : the Bulgarian Labour Market Perspective.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

205. Darkiewicz, Grzegorz (28/02/2005) *Value-at-risk in Insurance and Finance : the Comonotonicity Approach.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

206. De Moor, Lieven (20/05/2005) *The Structure of International Stock Returns : Size, Country and Sector Effects in Capital Asset Pricing.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

207. Everaert, Greetje (27/06/2005) *Soft Budget Constraints and Trade Policies : The Role of Institutional and External Constraints.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

208. Simon, Steven (06/07/2005) *The Modeling and Valuation of complex Derivatives : the Impact of the Choice of the term structure model.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

209. Moonen, Linda (23/09/2005) *Algorithms for some graph-theoretical optimization problems.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

210. Coucke, Kristien (21/09/2005) *Firm and industry adjustment under de-industrialisation and globalization of the Belgian economy.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

211. Decamps, Marc (21/10/2005) *Some actuarial and financial applications of generalized diffusion processes.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

212. Kim, Helena (29/11/2005) *Escalation games: an instrument to analyze conflicts. The strategic approach to the bargaining problem.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2005.

213. Germenji, Etleva (06/01/2006) *Essays on the economics of emigration from Albania.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

214. Belien, Jeroen (18/01/2006) *Exact and heuristic methodologies for scheduling in hospitals: problems, formulations and algorithms.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

215. Joossens, Kristel (10/02/2006) *Robust discriminant analysis.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

216. Vranken Liesbet (13/02/2006) *Land markets and production efficiency in transition economies.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

217. Vansteenkiste Isabel (22/02/2006) *Essays on non-linear modelling in international macroeconomics.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

218. Wuyts Gunther (31/03/2006) *Essays on the liquidity of financial markets.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

219. De Blander Rembert (28/04/2006) *Essays on Endogeneity and Parameter Heterogeneity in Cross-Section and Panel Data.* Leuven, K.U.Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.