FOUR ESSAYS ON DYNAMIC PANEL MODELS

by

Charles Saunders

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics

Carleton University

Ottawa, Ontario

© 2015, Charles Saunders

ABSTRACT

Dynamic panel data models can suffer greatly from incidental parameter bias due to correlation between past realizations of the data and the unobserved heterogeneity, and this bias is a function of included regressors.

This paper uses simulation-based methods that require explicit models and sets of assumptions to obtain consistent point estimates and exact confidence sets. A parametric discontinuous starting value is assumed for simulated series that jointly allows for stationary and unit root processes, where only the stationary case was considered in Gouriéroux, Phillips, and Yu (2010). This discontinuous assumption leads to least squared dummy variable (LSDV) estimator that are nuisance parameter free and location-scale invariant. These properties are conferred to the indirect inference objective function (IIOF) used to obtain bias-corrected estimates.

Discontinuities are problematic for traditional asymptotic methods of constructing confidence sets. To account for this the indirect confidence set inference method is introduced, which uses a second round of Monte Carlo simulations [Dufour (2006)] to calibrate the distribution of the IIOF. The confidence set is constructed with test inversion, so the parameters are set to known values, the model is tested at that point, and all points that fail to reject the null hypothesis are in the confidence set. The confidence set is exact and level correct, since the IIOF is pivotal and both simulation rounds are exchangeable under the null.

Adding regressors into panel data models can distort estimates, as this paper demonstrates with respect to the X-differencing method of Han, Phillips, and Sul (2014) with regressors. By introducing a model augmentation approach, the influence of regressors are corrected. The augmentation uses a projection of the regressors for

all time periods onto the model, giving the augmented-LSDV and the augmented-IIOF that are provably location-scale invariant and nuisance parameter free.

The framework presented is robust to non-Gaussian errors, which is demonstrated with bank cost data where the dynamic technical efficiency framework allows for skew-Normal errors. Simulation studies support all theoretical results, with confidence sets that are level correct with good coverage properties, even for asymmetric confidence regions.

ACKNOWLEDGMENTS

A special thank you to my supervisor Lynda Khalaf, that without your patience, support, and knowledge I would likely not have been able to advance as quickly and cleanly as I have. I would also like to thank Russell Davidson (McGill) for pointing out that the initial observation could be random. I would like to thank Jeffery Wooldridge, for our brief discussion in Budapest, and for indicating that he had given a block-diagonal Mundlak device some consideration in response to query at a conference a few years prior.

I would also like to thank my comrades in the PhD program, but specifically our microeconomics, macroeconomics and econometrics study groups: Sarah Mo-han, Duangsuda (Neat) Sopchokchai, Anand Acharya, Bogdan Urban, and Nyamekye Asare. A thank you to some others who have helped along the way: Marie-Hélène Felt, Zhenjiang Lin, Samira Hasanzadeh, Chaohui Lu, and Susan Bennet.

Thank you Abeer Reza for continuously pushing and cajoling me to start a PhD, with the confidence that I would really enjoy it - I have.

Last but not least, a huge thank you to Kevin McEvoy for helping me relax and laugh at the lighter side of life.

To my wife and boys.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
LIST OF TABLES	ix
LIST OF FIGURES	x
ABBREVIATIONS	xi
1 Introduction	1
1.1 Dynamic Panel Models	1
1.2 Summary	4
2 Confidence Intervals in Autoregressive Panels, Invariance, and Indirect Inference	7
2.1 Introduction	7
2.2 Framework	11
2.2.1 LSDV and Bias	12
2.2.2 Initial Observation	13
2.3 Invariance	17
2.4 Discontinuity, Indirect Inference, and Confidence Set	19
2.5 Simulation Study	23
2.5.1 Recursive-Design Wild-Bootstrap Confidence Set	24
2.5.2 Simulation Results	26
2.6 Concluding Remarks	37
3 Dynamic Panels, Exogenous Regressors, and X-Differencing	39
3.1 Introduction	39
3.2 Methods	40
3.3 Simulations	43
3.4 Final Remarks	45

	Page
4 On Autoregressive Panels with Exogenous Regressors: Invariance, Indirect Inference, and Confidence Sets	46
4.1 Introduction	46
4.2 Framework	50
4.2.1 LSDV with Exogenous Regressors	51
4.2.2 A Comment on Kiviet and van den Doel (1994) Invariance	54
4.3 Augmented-LSDV	56
4.3.1 Projection Matrix Properties	61
4.3.2 Recovery of β	63
4.4 Augmented-LSDV Invariance	65
4.5 Indirect Inference Estimation: ϕ and β	66
4.5.1 Bias Correction: Monte Carlo Evidence	70
4.6 Confidence Set for ϕ	72
4.7 Simulation Study	75
4.8 Concluding Remarks	76
4.9 Tables	78
5 Dynamic Technical Efficiency	83
5.1 Introduction	83
5.2 Methods	84
5.3 Unbiased Estimators and Confidence Set	87
5.4 Simulations	89
5.5 Banks, Productivity, and Dynamics	91
5.6 Final Remarks	93
6 Concluding Remarks	94
LIST OF REFERENCES	95
A Proofs for Chapter 2	99
A.1 Proof of Theorem 2.3.1	99
A.2 Corollary Proofs	101

	Page
B Proofs for Chapter 4	103
B.1 Proof of Theorem 4.4.1	104
B.2 Corollary Proofs	105

LIST OF TABLES

Table	Page
2.1 LSDV Bias Sensitivity to σ_γ	16
2.2 Rejection Frequency Sensitivity to Choice of σ_γ	27
3.1 X-differencing Parameter Estimation Bias	44
4.1 Parameter Estimate Bias for a Panel-ARX(1,1) Model	71
4.2 Rejection Frequency: $\phi = 0$ at 5% Significance Level	78
4.3 Rejection Frequency: $\phi = 0.6$ at 5% Significance Level	79
4.4 Rejection Frequency: $\phi = 0.9$ at 5% Significance Level	80
4.5 Rejection Frequency: $\phi = 0.999$ at 5% Significance Level	80
4.6 Rejection Frequency: $\phi = 1$ at 5% Significance Level	81
4.7 Rejection Frequency: $\phi = 0.6$, $N = 200$, and $T = 5$ at 5% Significance Level	82
5.1 Estimation Bias for a Dynamic Stochastic Frontier Model	89
5.2 ICSI Rejection Frequency for Lagged Dependent Parameter	90
5.3 ICSI Rejection Frequency for Shape Parameter	90
5.4 Dynamic Translog Model of Commercial Banks	92

LIST OF FIGURES

Figure	Page
2.1 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=100, T=5$	28
2.2 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=100, T=20$	28
2.3 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=300, T=5$	29
2.4 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=300, T=20$	29
2.5 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=100, T=5$	30
2.6 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=100, T=20$	30
2.7 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=300, T=5$	31
2.8 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=300, T=20$	31
2.9 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=100, T=5$	32
2.10 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=100, T=20$	32
2.11 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=300, T=5$	33
2.12 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=300, T=20$	33
2.13 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=5$	34
2.14 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=20$	34
2.15 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=5$	35
2.16 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=20$	35
2.17 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=5$	36
2.18 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=20$	36
2.19 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=5$	37
2.20 Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=20$	37

ABBREVIATIONS

AR	Autoregressive model
FWL	Frisch-Waugh-Lovell
GLS	Generalized least squares
GMM	Generalized method of moments
ICSI	Indirect confidence set inference
II	Indirect inference
IIE	Indirect inference estimator
IIOF	Indirect inference objective function
IV	Instrumental variable
LSDV	Least squared dummy variable estimator
MC	Monte Carlo
MLE	Maximum likelihood estimator
MMC	maximized Monte Carlo
OLS	Ordinary least squares
panel-AR	Panel autoregressive model
panel-ARX	Panel autoregressive model with exogenous regressors
RDWB	Recursive design wild bootstrap

1. INTRODUCTION

The focus of this work is on panel data models under the assumption that an observation in the current period is a function of past observations, where these models are defined as dynamic panel data models.

1.1 Dynamic Panel Models

The data structure of a dynamic panel data model is, at a minimum, a dual-indexing of observations where the first index is the cross-section and the second is time. In these data, each cross-section is observed for at least three time periods. The dynamic panel model for the dependent variable, $y_{i,t}$, is given by,

$$y_{i,t} = \phi y_{i,t-1} + x'_{i,t} \beta + (1 - \phi) \mu_i + \epsilon_{i,t}, \quad (1.1)$$

where the cross-section index is $i = 1, \dots, N$, the time index is $t = 1, \dots, T$, exogenous regressors are represented by $x_{i,t}$, μ_i is the unobserved heterogeneity of each cross-section, and $\epsilon_{i,t}$ are zero mean independently and identically distributed idiosyncratic errors. The model parameters are $\theta = (\phi, \beta)'$.

In panel data without dynamics (where $\phi = 0$), consistent estimation of the model parameters depends on the following moment conditions:

- strict exogeneity: $E[\epsilon_{i,t} | x_{i,1}, \dots, x_{i,T}, \mu_i] = 0$, and
- finite and homoscedastic variance: $V[\epsilon_i | x_i, \mu_i] = \sigma_\epsilon^2 I_N$ and $0 < \sigma_\epsilon^2 < \infty$.

This would be considered the most restrictive set of assumptions that are given and may be relaxed depending on both the model assumption and the estimation method.

However, for models with dynamics ($\phi \neq 0$), the first moment condition does not hold, so sequential exogeneity must be assumed, specifically

$$E[\epsilon_{i,t} | y_{i,1}, \dots, y_{i,t-1}, x_{i,1}, \dots, x_{i,T}, \mu_i] = 0.$$

The introduction of dynamics also results in a non-zero correlation between the lagged dependent series and the unobserved heterogeneity, $cov[\mu_i, y_{i,t-1}] \neq 0$, so parameter estimates suffer from an incidental parameter bias.

This incidental parameter bias was first discussed in Neyman and Scott (1948), and later Nickell (1981) derived the analytical bias for a model without regressors as a function of the true lagged dependent parameter (ϕ) and the number of time periods (T). The analytical bias for a dynamic panel data model with regressors is not easily derived since the bias is a function of ϕ , T , β , the number of exogenous regressors, and the variance ratio of σ_ϵ and Σ_X (variance-covariance matrix of the regressors). Phillips and Sul (2007) discuss the bias when regressors are present in the context of the least squares dummy variable estimation.

To correct for the incidental parameter bias, several instrumental variable (IV) and generalized method of moment (GMM) approaches have been put forth including those of Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Blundell and Bond (1998), and Bun and Kleibergen (2013). These IV and GMM methods each rely on some combination of the following moment conditions:

1. $E[y_{i,s}(\Delta y_{i,t} - \phi \Delta y_{i,t-1})] = 0$ for $t = 3, \dots, T$ and $s = 1, \dots, t-2$,
2. $E[\Delta y_{i,t-1}(y_{i,t} - \phi y_{i,t-1})] = 0$ for $t = 3, \dots, T$,
3. $E[\Delta y_{i,t-1} \mu_i] = 0$, and
4. $E[(y_{i,t} - \phi y_{i,t-1})(\Delta y_{i,t-1} - \phi \Delta y_{i,t-2})] = 0$ for $t = 4, \dots, T$.

The number of available moment conditions increases for models that include exogenous regressors.

The IV and GMM approaches can suffer from a weak instrument¹ problem especially as the lagged dependent parameter approaches unity. Also instrument proliferation poses a serious problem in GMM estimation since the number of valid instruments increases at a rate proportional to T^2 , so it is possible that the number of valid instruments can outstrip the number of observations even for relatively low values of T . The current wisdom is to exclude instruments (i.e., limits on lags) or apply shrinkage methods to the instrument set, for example LASSO (Lu and Su (2015)), factor analysis (Trapani (2015)), and the approach of Roodman (2009).

Another method of obtaining consistent estimates is a well defined bias-correction term, that is a function of the LSDV estimate from the data and the number of time periods. The bias-correction term is derived under either Large- N and Fixed- T conditions, in the case of Kiviet (1995) and Bun and Carree (2005), or in a Large- N and Large- T setting, for Hahn and Kuersteiner (2002) and Hahn and Moon (2006). Monte Carlo experiments show that these estimators can outperform the IV/GMM alternatives, but this result is far from universal.

A third method is that of Gouriéroux, Phillips, and Yu (2010), who extend the indirect inference estimator (IIE) to the LSDV estimate of a panel-AR(1) model. IIE is a simulation-based method that generates a set of simulated dynamic panel data under the null model and for a given set of parameters. The intuition of the IIE is that the bias term of the original data is a function of the true lagged dependent parameter and the number of time periods (T), as shown in Nickell (1981). Simulating the bias requires an assumption about the true value of the parameter and some distributional assumptions regarding the errors. The difference between the data bias and the simulated bias can be minimized, resulting in a consistent and unbiased estimate of the true parameter, under the null.

Some recent advances focus on constructing the moment conditions of the model such that the incidental parameter bias is controlled and estimates are unbiased and consistent. A new GMM approach is that of Han and Phillips (2010) which constructs

---

¹An instrument is classified as weak if it is orthogonal to endogenous variable that is being instrumented.

the moment conditions by augmenting with leads and lags. Another novel approach is the X-differencing of Han, Phillips, and Sul (2014), which converts the model into a forward looking model then differences it from a suitably distant backward looking model. This review of methods is by no means exhaustive of the entire literature on dynamic panel data model estimation, but focuses on the main approaches used in empirical studies or recent advances.

1.2 Summary

This work starts at the indirect inference (II) methodology for dynamic panel models outlined in Gouriéroux, Phillips, and Yu (2010), which employs the least squared dummy variable (LSDV) estimator to attain consistent parameter estimates. As a simulation-based method, II must explicitly define the model and all assumptions required for simulation. Gouriéroux et al. (2010) impose stationarity, $|\phi| < 1$, assume that the idiosyncratic errors follow a normal distribution, and primarily an autoregressive panel of the first order without exogenous regressors.

Chapter 2 shows that their assumptions can be extended to the unit root, $\phi = 1$, case by assuming a discontinuous starting value for the first observation of a panel data series. The II approach is robust to this assumption allowing for bias-corrected estimates for $\phi = (-1, 1]$; however, the discontinuity at unity is problematic for confidence sets based on asymptotic limits. The LSDV estimator with the discontinuous initial observation is provably location-scale invariant, and this property is conferred on the indirect inference objective function (IIOF) ensuring that it is nuisance parameter free, or pivotal. The indirect confidence set inference (ICSI) method is an inversion approach that sets ϕ to a given value ϕ_0 , then tests the model at that point by adding a second layer of Monte Carlo simulations to the II framework. This second layer of simulations allows for the computation of the Monte Carlo p-value, under the null of $\phi = \phi_0$. The IIOF is the statistic for this inversion and since it is pivotal and

the Monte Carlo and II simulations are exchangeable under the null, the resulting confidence sets are exact and level correct.

The simulation study supports the theoretical findings with level correct rejection frequencies under the null. Rejection frequencies under the alternative increase as ϕ_0 deviates from the true value, and the method demonstrates no distortions for local-to-unity versus the unit root panel.

Chapter 3 examines the X-differencing approach of Han, Phillips, and Sul (2014) when exogenous regressors are included with respect to finite-sample bias. Han et al. (2014) show in a dynamic panel model without regressors that the difference of a lead-based model from a lag-based model removes the incidental parameter bias. Their paper addresses the inclusion of exogenous regressors, but I find that the X-differencing method is provably biased for dynamic panel models when exogenous regressors are present. The forward-looking error component is correlated with both the exogenous regressors from the forward-looking model, and the ‘lag-lead’ dependent series, thereby violating the moment conditions of the model. This chapter demonstrates that the exogenous regressors need to be considered carefully in panel data models, since the usual exogeneity assumptions for time-series models may not always hold.

Chapter 4 introduces a model augmentation approach, described as a block-diagonal Chamberlain device, which removes the nuisance parameter properties associated with the exogenous regressor by projecting on all exogenous regressors for all time periods. This approach exploits the condition that the number of cross-sections is much larger than the number of time periods. The series resulting from the projection can be used in many of the existing bias correction methods. To illustrate this, the augmentation is applied to the LSDV estimator under both the II and ICSI methods. The augmented-LSDV and augmented-IIOF are location-scale invariant with no nuisance parameters, so are only functions of the lagged dependent parameter (ϕ). A method is presented to allow for the recovery of coefficients and inference on the exogenous regressors. The discontinuous initial observation assumption from Chapter

2 is updated to allow for exogenous regressors. The II estimates generated are bias corrected, and the confidence sets from the ICSI approach are exact and level correct.

Chapter 5 is an extension of the previous chapters to dynamic panels with non-Gaussian errors. These panels still suffer from the incidental parameter bias. Simulations show that an indirect inference estimation approach provides bias correction for the model and distribution parameters. The indirect confidence set inference method is size correct and exhibits good coverage properties even for asymmetric confidence regions. Bank cost data are examined under the proposed dynamic technical efficiency framework with evidence that a MLE approach could provide misleading implications.

Chapter 6 concludes with a discussion of potential extensions of the ICSI and the augmentation approach to other areas in econometrics.

2. CONFIDENCE INTERVALS IN AUTOREGRESSIVE PANELS, INVARIANCE, AND INDIRECT INFERENCE

Charles J. Saunders and Lynda Khalaf

2.1 Introduction

The least squares dummy variable (LSDV) estimate of an autoregressive panel is inconsistent¹, for a given number of time periods. The source of the bias is an incidental parameter problem, first identified by Neyman and Scott (1948), since the lagged dependent variable is correlated with the time-invariant component of the model error. Demeaning and first differences alone are ineffective at eliminating this bias.

A common method of bias correction is based in the instrumental variable (IV) and generalized method of moment (GMM) literature. The most popular of these estimators were introduced by Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), and Blundell and Bond (1998). The instrumental variables used are exogenous variables (if present) and estimator specific combinations of level and first difference lags of the dependent variable, and lags of the error estimates. These estimators generally rely on Large- N and Small- T for both consistency and good asymptotic properties. As shown in Bun and Kleibergen (2013) and the references therein, regularity conditions may hold only weakly, especially with Small- T settings, which may cause non-uniform convergence problems even when the number of cross-sections is large. Weak-instruments concerns complicate limiting theory, a problem not restricted, of course, to panel data models.

---

¹Inconsistent in the sense that the number of cross-sections tends to infinity for a fixed number of time periods.

Bun and Kleibergen (2013) show that identification of the parameters of two-step estimators have non-standard distributions in large samples, which is in line with the findings of Kruiniger (2009). Bun and Kleibergen (2013) also show that under weak instrument conditions, at or near the unit root boundary, then some estimators can identify the parameters, specifically: Newey and West (1987), Stock and Wright (2000) and Moreira (2009). For a more complete summary of weak instruments and dynamic panel data models, see Bun and Sarafidis (2015).

Another method of obtaining consistent estimates is a well defined bias-correction term, that is a function of the LSDV estimate from the data and the number of time periods. The bias-correction term is derived under either Large- N and Fixed- T conditions, in the case of Kiviet (1995) and Bun and Carree (2005), or in a Large- N and Large- T setting, for Hahn and Kuersteiner (2002) and Hahn and Moon (2006). In Monte Carlo experiments, these estimators have been shown to outperform the GMM/IV alternatives but not for all simulation settings.

A third method is that of Gouriéroux, Phillips, and Yu (2010), extending the indirect inference estimator (IIE) to the LSDV estimate of a panel-AR(1) model. IIE is a simulation-based method that generates a set of dynamic panel data under the null model and for a given set of parameters. The intuition of the IIE is that the bias term of the original data is a function of the true lagged dependent parameter and the number of time periods (T), as shown in Nickell (1981). Simulating the bias requires an assumption about the true value of the parameter and some distributional assumptions regarding the errors. The difference between the data bias and the simulated bias can be minimized, resulting in a consistent and unbiased estimate of the true parameter, under the null.

The dynamic panel literature has focused primarily on consistent estimators. In most cases, these estimators impose stationarity to ensure appropriate asymptotic properties. Unit root panel models have, for the most part, followed a separate research thread and crossing the stationary thread on occasion. Even more generally, outside the panel context, a method that works at or close to the boundary may fail

elsewhere. Phillips (2014a) shows that constructing confidence intervals by inverting an alternative model can result in not only poor coverage properties but even complete lack of coverage. The asymptotic properties of stationary, local-to-unity, and unit root models differ greatly, which is a serious impediment to developing a unified asymptotic theory.

Motivated by this chapter makes three contributions. Our first contribution is the introduction of a discontinuous panel-AR(1) framework suitable for joint consideration of stationary, local-to-unity, and unit root panels. The initial observation of the series interferes with the LSDV estimate of the lagged dependent parameter, hence it is a nuisance parameter. We demonstrate that ignoring the nuisance parameter causes serious size distortions. To solve this problem, we propose a discontinuous functional specification of the initial observation in terms of the lagged dependent parameter as well as other hyper-parameters, specifically unobserved heterogeneity and the variance of the time-varying error. Our initial observation specification respects current practice in the literature and leads to an objective function that is provably location-scale invariant with analytical proofs provided. Our proposed approach demonstrates that a discontinuity that is an impediment to unifying asymptotic theory can be an asset in a finite-sample context.

Invariance proofs in the dynamic panel context are limited. Kiviet and van den Doel (1994) prove invariance of a LSDV estimator that has been augmented with an additional lag of exogenous regressors. Their augmented-LSDV estimator relies on exogenous regressors for invariance and four or more time periods.

Our second contribution extends the indirect inference estimator in Gouriéroux, Phillips, and Yu (2010) to jointly consider stationary and unit root panels, through the inclusion of our discontinuous framework. Standard objective functions are ill-behaved at or close to unity since the long-run variance of the series is not finite. Our modification corrects this problem without costing power. In contrast to point estimation through minimization, we focus on confidence set estimation by inverting the objective function.

Formally, our proposed method treats the indirect inference objective function as the test statistic which we invert rather than minimize. Inversion of our test statistic involves collecting the lagged dependent parameter values that are not rejected at a given level. We generate two sets of simulated series under our joint null, which is comprised of a given lagged dependent parameter value and our discontinuous initial observation function. The first set of simulated series are used to define out test statistic that is computed as the objective function realization from the data under the null. The test statistic is standardized by the second set of simulated series and a simulation-based p-value is obtained by Monte Carlo methods [refer to Dufour (2006), Dufour and Valéry (2009) and Beaulieu, Dufour, and Khalaf (2007, 2013, 2014)]. We show that the Monte Carlo technique allows for discontinuity under the null and works to construct exact confidence sets, even though conventional methods could lead to highly inaccurate inference. Our proposed method is an interesting contribution to the theory of test inversion and simulation-based methods.

Our third contribution examines the properties of the indirect confidence set inference method through a simulation study. The indirect confidence set inference method exhibits level control and power, even under modest cross-sections and small number of time periods conditions. The level control and power properties of the method are retained for stationary and unit root autoregressive panels. For the stationary panel-AR(1) simulations, as a means of comparison we examine the coverage of the bootstrap method of Gonçalves and Kaffo (2014).

The paper is structured as follows, section 2.2 details our panel-AR(1) model, assumptions, parameter estimation, invariance proofs and bias. Section 2.4 presents the indirect inference method to the panel-AR(1) model, and extends it to allow for the assumption of discontinuous initial value function. Also in this section, we introduce the indirect confidence set inference method to the panel-AR(1) model to obtain an exact confidence set for the parameter. In section 2.5, we present the design and results of our simulation study for our method and find level control and

good power properties. Section 2.6 summarizes the major findings and indicates some potential extensions to non-Gaussian disturbances.

2.2 Framework

The general form of a panel-AR(1) model is given as,

$$\begin{aligned} y_{i,t} &= (1 - \phi) \mu_i + \phi y_{i,t-1} + \epsilon_{i,t} \\ i &= 1, \dots, N \\ t &= 1, \dots, T, \end{aligned} \tag{2.1}$$

where i indicates the cross-section of the panel, t is the time index of the observations, and $y_{i,t}$ is the dependent variable. Alternative representations of this model have been implemented in the literature and are discussed below. The representation given above is robust to roots of the lag polynomial that lie outside or on the unit circle. The $\epsilon_{i,t}$ are idiosyncratic errors that follow a distribution centered at zero and allows for independent draws. The time-invariant μ_i terms can either be non-random or random disturbances with the first two moments defined as μ and σ_μ^2 . The existence of these moments does not intervene in our proposed approach presented below since our theorems do not rely on asymptotics but instead a finite-sample approach. It is convention that $\epsilon_{i,t}$ and μ_i are contemporaneously uncorrelated,

$$E[\epsilon_{i,t} | \mu_j] = 0 \quad \forall i, j.$$

Common in the literature is the assumption that both errors follow a Gaussian distribution, but clearly the framework is robust to most alternative distributional assumptions. Chapter 5 introduces this framework to non-Gaussian errors, where the idiosyncratic errors follow a skew-normal distribution.

2.2.1 LSDV and Bias

The panel-AR(1) model presented in equation 2.1 in demeaned form is

$$(y_{i,t} - y_{i,\bullet}) = \phi(y_{i,t-1} - y_{i,\bullet-1}) + (\epsilon_{i,t} - \epsilon_{i,\bullet})$$

where the sample means are given by

$$y_{i,\bullet} = \frac{1}{T-1} \sum_{t=2}^T y_{i,t}, \quad y_{i,\bullet-1} = \frac{1}{T-1} \sum_{t=1}^{T-1} y_{i,t}, \quad \text{and} \quad \epsilon_{i,\bullet} = \frac{1}{T-1} \sum_{t=2}^T \epsilon_{i,t}.$$

A condition of the LSDV estimator is that minimum number of time periods is defined by $T \geq 3$. We set a distinction here, so for the models presented the dependent variable is observed for $t \geq 1$, and not observed for $t \leq 0$, so LSDV estimate is based on the observed series.² To simplify the notation, the demeaned series will be denoted with an asterisk (*). So the estimated equation is

$$y_{i,t}^* = y_{i,t-1}^* \phi + \epsilon_{i,t}^*$$

and through appropriate stacking we obtain the matrix form,

$$y^* = y_{-1}^* \phi + \epsilon^*. \quad (2.2)$$

The LSDV estimates of ϕ for equation 2.2 gives

$$\hat{\phi} = (y_{-1}^{*'} y_{-1}^*)^{-1} y_{-1}^{*'} y^* = \phi + (y_{-1}^{*'} y_{-1}^*)^{-1} y_{-1}^{*'} \epsilon^*. \quad (2.3)$$

The estimate of ϕ is biased due to $E[y_{-1}^{*'} \epsilon^*] \neq 0$ for fixed-T, and this bias is analytically derived in Nickell (1981).

---

²An alternative approach divides the observed series into modeled and initial observations with notionally equivalent implications.

2.2.2 Initial Observation

The unobserved lead values ($t \leq 0$) affect parameter estimation in dynamic models since their influence persists. The general assumption is that the series was initiated long enough in the distant past so that the effects of the initial observation are negligible.

A historical approach was to simulate a series that was longer than required then retain only the most recent observations, an example of this approach is found in the Monte Carlo study of Arellano and Bond (1991). However, the number of truncated observations depends on the model parameters, as parameter settings approach nonstationarity the number of required truncated observations increases.

Alternatively, the initial observation can be specified parametrically, which relaxes the need to generate and truncate additional lead observations. Our parametric specification is stated as assumption 1 and is a general in the sense that it encompasses the majority of parametric specifications used in the literature. The dispersion of the initial observation is determined by σ_γ .

Assumption 1 (Initial Values for Simulated Series) *The initial value of the series for any i is defined by*

$$y_{i,0} = \mu_i + \sigma_\gamma \epsilon_{i,0}$$

where $\sigma_\gamma \geq 0$, and $\epsilon_{i,0}$ follows the same distribution as $\epsilon_{i,t}$ and $E[\epsilon_{i,0}\epsilon_{i,t}] = 0$ for all t .

In finite sample and standard time series approaches it is common to condition on the initial observation by setting the dispersion of the initial observation to zero.

Assumption 2 (Conditional σ_γ) *Let the initial value of the series for any i under assumption 1 be a function of $\sigma_\gamma(\phi)$, which is defined as*

$$\sigma_\gamma := \sigma_\gamma(\phi) = 0 \text{ for all } \phi.$$

Assumption 2 means that the initial observation is equal to the unobserved heterogeneity of the series, $y_{i,0}(\phi) = \mu_i$. In the dynamic panel context, Gonçalves and Kaffo (2014) impose an initial value setting under no dispersion,

$$\begin{aligned} y_{i,t} &= a_i + \phi y_{i,t-1} + \epsilon_{i,t} \text{ and} \\ y_{i,0} &= \frac{a_i}{1 - \phi}. \end{aligned}$$

This setting can be transformed into equation 2.1 by substituting the following identity, $a_i = (1 - \phi)\mu_i$.

The importance of initial observation assumptions is examined numerically in the GMM framework by Hahn (1999), finding evidence in favour of imposing stationarity. To account for the unit root, we introduce a discontinuous null setting for σ_γ that is a function of ϕ .

Assumption 3 (Discontinuous σ_γ) *Let the initial value of the series for any i under assumption 1 be a function of $\sigma_\gamma(\phi)$, which is defined as*

$$\sigma_\gamma := \sigma_\gamma(\phi) = \begin{cases} \frac{1}{\sqrt{(1-\phi^2)}} & \text{if } |\phi| < 1, \\ 0 & \text{if } \phi = 1. \end{cases}$$

Under assumptions 1 and 3, the final unobserved value is discontinuous in ϕ , specifically

$$\begin{aligned} y_{i,0}(\phi) &= \mu_i + \sigma_\gamma(\phi)\epsilon_{i,0} \\ &= \begin{cases} \mu_i + \frac{1}{\sqrt{(1-\phi^2)}}\epsilon_{i,0} & \text{if } |\phi| < 1, \\ \mu_i & \text{if } \phi = 1. \end{cases} \end{aligned}$$

As $\phi \rightarrow 1$ then the final unobserved value is ill-defined in a stationary setting, since $y_{i,0}(\phi) \rightarrow \infty$, affecting the parameter estimates unless a functional form is imposed that takes into account the discontinuity at unity. These problems at the

limit are avoided under assumption 3, because at unity the final unobserved value is only a function of the time-invariant error. Corollary 2, given below, justifies our approach since the LSDV estimate of ϕ for a unit root series is invariant to σ_γ . Under assumption 3, σ_γ is no longer a nuisance parameter since it is completely defined under the null.

Assumption 1 is a general specification that encompasses available final unobserved value settings, below we provide a brief survey of the data generation processes available in the literature. We show that these can be defined as special cases of assumptions 1 and 3, for stationary dynamic panels. Hahn and Moon (2006) use the following settings,

$$\begin{aligned} y_{i,t} &= (1 - \phi)\mu_i + \phi y_{i,t-1} + \epsilon_{i,t} \text{ and} \\ y_{i,0} &\sim \mathcal{N}\left(\mu_i, \frac{1}{1 - \phi^2}\right). \end{aligned}$$

Setting $\epsilon_{i,0} \sim \mathcal{N}(0, 1)$ demonstrates this as a special case of assumption 1.

Gouriéroux, Phillips, and Yu (2010), Bun and Carree (2005), and Hahn and Kuersteiner (2002) assume the following set of equations,

$$\begin{aligned} y_{i,t} &= a_i + \phi y_{i,t-1} + \epsilon_{i,t} \text{ and} \\ y_{i,0} &= \frac{a_i}{1 - \phi} + \frac{\epsilon_{i,0}}{\sqrt{1 - \phi^2}}, \end{aligned}$$

where the initial observation is defined by

$$y_{i,0} \sim \mathcal{N}\left(\frac{a_i}{1 - \phi}, \frac{\sigma_\epsilon^2}{1 - \phi^2}\right).$$

This framework can be transformed into our framework, $a_i = (1 - \phi)\mu_i$, and the initial value is obtained by assuming $\epsilon_{i,0} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and $\sigma_\gamma = (1 - \phi^2)^{-1/2}$. Kiviet and van den Doel (1994) impose this special case when no exogenous regressors are present.

The papers cited above post-date the analytical bias derived in Nickell (1981). Table 2.1 shows that the Nickell bias of the LSDV estimator is sensitive to assumptions on the magnitude of σ_γ . The effect of the final unobserved value does not get eliminated through demeaning, which is apparent by the alternative representation of the demeaned dependent series, see Lemma 1 in the next section. This additional term intervenes with the LSDV estimate of ϕ that is proportional to the square of term.

Table 2.1: LSDV Bias Sensitivity to σ_γ

ϕ	σ_γ					
	0	1	1.25	1.5	2	3
0	-0.2148	-0.2148	-0.2148	-0.2148	-0.2148	-0.2148
0.5	-0.4496	-0.4346	-0.4269	-0.4179	-0.3977	-0.3544
0.6	-0.5086	-0.4901	-0.4807	-0.4699	-0.4461	-0.3971
0.7	-0.5694	-0.5506	-0.5410	-0.5300	-0.5056	-0.4553
1	-0.7153	-0.7153	-0.7153	-0.7153	-0.7153	-0.7153

Based on 1000 Monte Carlo replications, $N = 200$, $T = 5$,
and standard deviation of the bias ranges from 0.0006 to 0.0009.

For $\phi = 0$ and $\phi = 1$ the value of σ_γ does not change the value of the LSDV estimates. For $0 < \phi < 1$, as the value of σ_γ increases the bias is reduced, so σ_γ is a nuisance parameter for the LSDV bias. The table above shows that in the local region of a given ϕ the bias is not unique when σ_γ is a free parameter. The initial observation can lead to large distortions in the LSDV estimator properties without a coherent specification of σ_γ .

In Section 2.5, we revisit the σ_γ in terms of a nuisance parameter problem where simulations indicate serious size distortions resulting from arbitrary selection of σ_γ including the assumption of $\sigma_\gamma = 0$.

2.3 Invariance

The recursion of the series is the building block for the proofs and corollaries, so it is stated here as Lemma 1.

Lemma 1 (Recursive Decomposition) *The demeaned series by recursive substitution are*

$$\begin{aligned}
 y_{i,t}^* &= y_{i,t} - y_{i,\bullet} = g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T), \text{ where} \\
 g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) &= \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r \right) \\
 &\quad + \left(\sigma_\gamma \epsilon_{i,0} \phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \phi^s \right). \tag{2.4}
 \end{aligned}$$

Let $g_{i,t} = g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T)$ and $g_{i,t-1} = g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t-1, T)$, and the vector representations are g and g_{-1} , respectively. Then, the LSDV estimate of ϕ is

$$\tilde{\phi} = (g'_{-1}g_{-1})^{-1}g'_{-1}g = \phi + (g'_{-1}g_{-1})^{-1}g'_{-1}\epsilon^*. \tag{2.5}$$

Under assumption 1 and 3 and Lemma 1, the parameter estimate from equation 2.3 is numerically equal to the estimate from equation 2.5, or $\tilde{\phi} = \hat{\phi}$. This implies that the distribution of $\hat{\phi}$ is completely determined by current, lagged and lead errors (ϵ), the true value of ϕ , σ_γ , and unobserved random disturbances ($\epsilon_{i,0}$).

In the context of this paper, an estimator is invariant when a known data transformation results in an unchanged parameter estimate. A scale transformation is when the data multiplied by a positive constant, and a location transformation is the addition of a non-zero constant to either the entire series or individual cross-sections. In the LSDV framework, we focus on invariance in terms of the lagged dependent parameter (ϕ) estimate.

Theorem 2.3.1 (Scale and Location Invariance) *Under assumption 1 and for a given σ_γ under assumption 3, the least-square dummy variable estimate of ϕ from a model defined by*

$$y_{i,t} = (1 - \phi)\mu_i + \phi y_{i,t-1} + \epsilon_{i,t},$$

is scale invariant in terms of σ_ϵ and location invariant in terms of μ_i .

Proof of Theorem 2.3.1 is provided in the Appendix.

The LSDV estimate of ϕ is a function of the true parameter value σ_γ and a set of hyper-parameters, specifically: σ_ϵ^2 , σ_μ^2 , and μ . By Theorem 2.3.1, the LSDV estimator is invariant to the hyper-parameters. By assumption 3, we define σ_γ as a discontinuous function of lagged dependent parameter. Although seemingly trivial, in our approach the initial observation is completely defined by the lagged dependent parameter and the hyper-parameters and as such the estimate of ϕ does not change.

Apparent by Lemma 1, the estimate of ϕ depends on σ_γ . As $\sigma_\gamma \rightarrow \infty$, the demeaned series is dominated by the initial observation so the LSDV estimate of ϕ will tend to the true value as the relative importance of the bias is reduced. This is a consequence of the final unobserved value being drawn from a different distribution than the observed series, so the additional term dominates the least squares approach.

Corollary 1 *Under assumptions 1 and 3, the LSDV estimate is invariant to affine transformations of the data.*

Corollary 1 states that $\tilde{\phi} = \hat{\phi}$, where $\tilde{\phi}$ is the LSDV estimate from affine transformed data ($\tilde{y}_{i,t} = \kappa y_{i,t} + \delta_i$ and $\kappa \neq 0$), and $\hat{\phi}$ is the LSDV estimate from the untransformed data ($y_{i,t}$).

Corollary 2 *Under assumption 1, the LSDV estimate is invariant to σ_γ for $\phi = 1$.*

The implication of Corollary 2 is that the LSDV estimate, $\hat{\phi}$, is not a function of σ_γ , so is unchanged for any value of σ_γ as shown in Table 2.1. Corollaries 1 and 2 are exploited from a test inversion perspective in our proposed method presented below.

2.4 Discontinuity, Indirect Inference, and Confidence Set

The indirect inference estimator (IIE) is a simulation-based method introduced by Gouriéroux, Monfort, and Renault (1993) and independently by Smith (1993) and has been primarily used for bias correction. Gouriéroux, Phillips, and Yu (2010) apply the IIE to the panel-AR(1) context and show that their approach provides asymptotically consistent estimates of ϕ . All simulation-based approaches must explicitly define all assumptions under the null, which in the case of IIE includes the an explicit assumption on the final unobserved value. We address this by implementing assumptions 1 and 3 into the null.

The LSDV estimate of ϕ is obtained from the data and denoted as $\hat{\phi}_o$. We propose the following modification by setting a joint null,

$$H_s : \phi = \phi_s \text{ and } \sigma_\gamma = \sigma_\gamma(\phi_s),$$

to a reasonable range of parameter values, $\phi_s \in \Phi = (-1, 1]$. This joint null does not lack generality, since it covers the specific settings defined in the literature as special cases. Under the joint null, we generate H simulated series,

$$\begin{aligned} y_{i,0,h} &= \sigma_\gamma(\phi_s) \epsilon_{i,0,h} \\ y_{i,t,h} &= \phi_s y_{i,t-1,h} + \epsilon_{i,t,h}. \end{aligned}$$

where $h = 1, \dots, H$, and $\epsilon_{i,0,h}$ and $\epsilon_{i,t,h}$ are independent draws from a given distribution. We set $\mu_{i,h} = 0$ for all i , which is suitable by our invariance theorem.

For each of the simulated series we compute the associated LSDV parameter estimate, denoted $\hat{\phi}_{s,h}$:

$$\hat{\phi}_{s,h} = (y_{-1,h}^{*'} y_{-1,h}^*)^{-1} y_{-1,h}^{*'} y_h^*.$$

Each simulated parameter estimate is conditional on the value of ϕ_s used in the simulation model and the draws of the time-varying disturbances. Under our joint null,

as shown above, $\hat{\phi}_{s,h}$ are location-scale invariant for all h , hence nuisance parameter free.

The indirect inference objective function is a distance measure, represented by $\|\cdot\|$, between the estimate of ϕ from the data ($\hat{\phi}_o$) and a binding function ($b_T(\hat{\Phi}_{s,H})$) based on the estimates from the simulated series, specifically

$$Q(\phi_s) = \left\| \hat{\phi}_o - b_T(\hat{\Phi}_{s,H}) \right\|.$$

Simulating under the null relaxes the requirements that the binding function is analytically tractable, invertible, or continuous. Defining the binding function as

$$b_T(\hat{\Phi}_{s,H}) = E[\hat{\phi}_{s,h}],$$

allows for consistent estimation as $H \rightarrow \infty$ by the empirical average from the simulated series,

$$b_T(\hat{\Phi}_{s,H}) = \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{s,h}.$$

The resulting indirect inference objective function is

$$Q(\phi_s) = \left\| \hat{\phi}_o - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{s,h} \right\|.$$

Gouriéroux, Phillips, and Yu (2010) applied a distance metric in a sandwich form with the identity matrix as the weighting matrix:³

$$Q(\phi_s) = \left(\hat{\phi}_o - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{s,h} \right)' \left(\hat{\phi}_o - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{s,h} \right). \quad (2.6)$$

---

³Alternatives include: the Euclidean norm, and the absolute value of the distance. The absolute value may have desirable properties if the distribution of $\hat{\phi}_{s,h}$ is heavily skewed (asymmetric). The identity matrix is appropriate for our setting because there is a single parameter and the LSDV estimations and indirect inference objective function is scale invariant.

Theorem 2.4.1 (Objective Function Location-Scale Invariant) *Under assumptions 1 and 3, and the joint null, defined as $H_s : \phi = \phi_s$ and $\sigma_\gamma = \sigma_\gamma(\phi_s)$, the distribution of $Q(\phi_s)$ is completely determined by ϕ_s and all the disturbances.*

Theorem 2.4.1 means that the indirect inference objective function is fully defined by the parameter estimates from the data and the simulated series. Indeed by Theorem 2.3.1, the parameter estimates and the indirect inference objective function are nuisance parameter free, hence the objective function is pivotal under the joint null.

The information content of the binding function on the true parameter affects the quality of the estimate. Weakly defined binding functions result in less efficient bounds on the confidence set, which can occur when nuisance parameters are present, or whether the binding function is locally or globally injective. By Theorem 2.4.1, the indirect inference objective function is globally injective under the joint null.

We next proceed from a confidence set perspective. Formally, we build the exact confidence set for ϕ , which is valid even for our discontinuous objective functions. Our approach inverts the indirect inference objective function, Equation 2.6, to compute the p -value of ϕ at a given parameter value. For a given value of ϕ , say $\phi_s \in (-1, 1]$, we compute the p -value based on the Monte Carlo methodology outlined in Dufour (2006). We build the confidence set by retaining the given values of ϕ with a p -value greater than our desired significance level. For our approach, the distribution of indirect inference objective function under the null need not be known, but only that it can be simulated.

Under the joint null, we generate M Monte Carlo replicated series and respective LSDV estimate,

$$\begin{aligned} y_{i,0,m|s} &= \sigma_\gamma(\phi_s)\epsilon_{i,0,m} \\ y_{i,t,m|s} &= \phi_s y_{i,t-1,m|s} + \epsilon_{i,t,m} \\ \hat{\phi}_{s,m} &= (y_{-1,m}^{*'} y_{-1,m}^*)^{-1} y_{-1,m}^{*'} y_m^*, \end{aligned}$$

where $\epsilon_{i,0,m}$ and $\epsilon_{i,t,m}$ are drawn from the same distribution as $\epsilon_{i,0,h}$ and $\epsilon_{i,t,h}$, and $\mu_{i,m} = 0$.

The objective function for each Monte Carlo series is

$$Q(\phi_{m|s}) = \left(\hat{\phi}_{s,m} - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right)' \left(\hat{\phi}_{s,m} - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right),$$

which are collected to build a set of M Monte Carlo objective function realizations, denoted as $Q_M(\phi_s)$. The set $Q_M(\phi_s)$ is used to calibrate the objective function realization from the data under the null.

Under the null, the parameter settings and assumptions on the disturbances are identical for the Monte Carlo replications and simulated series. All parameter estimates from these series will be independent and identically distributed, hence any permutation of the combined set can be used to construct $\frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s}$ and each $Q(\phi_{m|s})$.

To compute the p -value under the null, we extend the method of Dufour (2006) to our statistic. We begin by counting the number of elements in $Q_M(\phi_s)$ that exceed $Q(\phi_{o|s})$, via a boolean check,

$$\text{count}(Q(\phi_{o|s})) = \sum_{m=1}^M \mathbf{1}(Q(\phi_{o|s}) \geq Q(\phi_{m|s}))$$

where $\mathbf{1}(\cdot)$ is one when the statement is true and zero otherwise. The p -value is computed as

$$\text{p-value}(Q(\phi_{o|s})) = \frac{1 + \text{count}(Q(\phi_{o|s}))}{1 + M}.$$

The confidence interval for ϕ at significance level α is constructed by collecting all the values of ϕ that satisfy the right-tailed condition,

$$CI(\phi|\alpha) = \{ \phi_s \mid \text{p-value}(Q(\phi_{o|s})) > \alpha \}.$$

The estimate of ϕ from the data and simulations are free of nuisance parameters, and conditional on a given N and T . By Theorem 2.4.1, the indirect inference objective functions are pivotal so the confidence interval will be exact. Since the simulated series and Monte Carlo replications are exchangeable under the joint null, then the confidence interval will have level control, in the sense of

$$Pr(\phi \in CI(\phi|\alpha)) = 1 - \alpha.$$

Alternative formulations of IIE have been suggested that include score-based by Gallant and Tauchen (1996), likelihood-distance-based by Smith (1993), and simulated GMM (refer to McFadden (1989)). Each of the three IIE methods are optimal under certain conditions, although the use of the *sample mean* is the least restrictive but most computationally intensive. With only a single parameter and reasonable values for H and M , the computation speed is negligible.

2.5 Simulation Study

We use a simulation study to examine the rejection frequency properties of the indirect confidence set inference method. Our method will be size correct under the null by construction, and the rejection frequency under the alternative (power) is of interest for ϕ coverage.

The LSDV estimations are invariant to both location and scale, under assumptions 1 and 3, so our simulation design needs specify ϕ , and the distribution of ϵ used for draws for the H simulated and M Monte Carlo simulated data.

For our tests we set parameter values that are non-dynamic, stationary, local-to-unity, and unit root, $\phi = \{0.0, 0.6, 0.9, 0.999, 1\}$. We build the confidence set under the joint null, outlined in the previous section, where $\phi_g \in (-1, 1]$. As a simulation study, we can limit our rejection frequencies of interest to a set of ϕ_g from 0.05 steps for over the closed interval $[-0.95, 0.95]$, and the points 0.999 and 1.

The number of time periods examined are $T = \{5, 10, 20\}$, and the range of the number of cross-sections are $N = \{100, 200, 300\}$. While considered modest, these combinations are representative of dynamic panel settings that are known to exhibit poor properties, also these combinations were used in Gouriéroux, Phillips, and Yu (2010) allowing for comparison. To compute our p-value, our approach sets $H = 20$ for the simulated series and $M = 79$ simulated Monte Carlo series. The rejection frequencies are based on 1000 replications, and computed as the number times ϕ_g is rejected at a given significance level (5%) divided by 1000.

The coverage of our proposed method is compared with the recursive design wild bootstrap of Gonçalves and Kaffo (2014) with our implementation of their procedure outlined below. It is important to note that the bootstrap of Gonçalves and Kaffo (2014) imposes the assumption that $\sigma_\gamma = 0$, so we remain true to their settings in our implementation.

The conventional IIE asymptotic standard errors are not valid under the joint null, so coverage of these confidence intervals are not examined. This is a consequence of the breaks in the asymptotic distribution of the LSDV estimate, specifically stationary (normally distributed), local-to-unity (Ornstein-Uhlenbeck Process), and unit root (functions of Brownian motion).

2.5.1 Recursive-Design Wild-Bootstrap Confidence Set

Gonçalves and Kaffo (2014) introduce three distinct bootstrap approaches, but only the recursive-design residual-based wild bootstrap (RDWB) method provides good bias control, in the context of a panel-AR(1) model. Their method employs the Hahn and Kuersteiner (2002) correction term

$$\hat{\phi} = \hat{\phi}_o + \frac{1}{T}(1 + \hat{\phi}_o) \quad (2.7)$$

where $\hat{\phi}_o$ is the LSDV estimate from the data. The Hahn and Kuersteiner (2002) bias-correction method has only been derived for the panel-AR(1) model, and is not appropriate for higher order autoregressive panels.

The Hahn and Kuersteiner (2002) estimate of ϕ is used to compute the estimation errors,

$$\hat{\epsilon}_{i,t} = y_{i,t} - \hat{\mu}_i - \hat{\phi}y_{i,t-1},$$

for $t = 2, \dots, T$ and for $t = 1$

$$\hat{\epsilon}_{i,1} = y_{i,1} - \hat{\mu}_i - \hat{\phi}y_{i,0},$$

where

$$y_{i,0} = \frac{\hat{\mu}_i}{1 - \hat{\phi}}.$$

It is important to note that the initial observation, $y_{i,0}$, restricts the model to $|\phi| < 1$ under the null. As such, the RDWB method is not examined at the boundary, $\phi = 1$, in our simulation study.

The wild-bootstrap errors are generated as

$$\tilde{\epsilon}_{i,t} = \hat{\epsilon}_{i,t} * \nu_{i,t} \text{ where } \nu_{i,t} \sim i.i.d.(0, 1).$$

In the Monte Carlo study, we use 999 independent draws of $\nu_{i,t}$ from the Rademacher distribution. The errors are used to recursively generate a set of simulated version of the dependent variable,

$$\tilde{y}_{i,t} = \hat{\mu}_i + \hat{\phi}\tilde{y}_{i,t-1} + \tilde{\epsilon}_{i,t}.$$

Each of the bootstrapped samples are used to estimate ϕ using the bias-correction Equation 2.7. The set of estimates are then sorted to obtain the bootstrap confidence limits of ϕ at the desired significance level.

2.5.2 Simulation Results

Figures 2.1 to 2.20 show the rejection frequency for the panel-AR(1) cases of the indirect confidence set inference (ICSI) method (red line) and the recursive-design wild bootstrap (RDWB) method (blue line).⁴ The ICSI method has good power properties along with level control, as shown in the figures. The lowest rejection frequency coincides with the data generating process value of ϕ . As the magnitude of the difference between the ϕ_s and DGP value increases, the rejection frequency rises, indicating power. As T and N increase, the rejection frequencies under the alternative for the ICSI method increase, while retaining level control under the null.

For $T = 5$, the RDWB method is extremely oversized for all values of ϕ . The RDWB method uses the Hahn and Kuersteiner bias-correction method, which relies on Large- T . The rejection frequency improves for $T = 20$ but still well above the significance level. An increase in N , as expected, does not provide any significant improvements to size, but power would be improved if the method was correctly centered.

---

⁴Rejection frequency tables under the null and alternatives for ICSI and RDWB are available upon request.

Table 2.2: Rejection Frequency Sensitivity to Choice of σ_γ

ϕ	σ_γ					
	0	1	1.25	1.5	2	3
0.30	0.999	0.999	0.999	0.999	0.994	0.965
0.35	0.988	0.982	0.975	0.966	0.925	0.714
0.40	0.939	0.896	0.870	0.815	0.660	0.301
0.45	0.794	0.655	0.591	0.507	0.314	<i>0.064</i>
0.50	0.557	0.374	0.297	0.206	<i>0.080</i>	0.165
0.55	0.309	0.155	0.092	<i>0.055</i>	<i>0.060</i>	0.551
0.60	0.129	<i>0.044</i>	0.048	<i>0.059</i>	0.202	0.830
0.65	0.039	<i>0.051</i>	0.091	0.167	0.489	0.958
0.07	<i>0.046</i>	0.146	0.238	0.393	0.709	0.990
0.75	0.153	0.353	0.470	0.604	0.865	0.999
0.80	0.410	0.592	0.680	0.784	0.935	1.000

Indirect Confidence Set Inference method based on 1000 Monte Carlo replications, $N = 200$, $T = 5$, $\phi = 0.6$ and 5% significance level.

Table 2.2 shows the rejection frequency distortions that can occur from an arbitrary choice of σ_γ . A series that initiated in the distant past with $\phi = 0.6$ will have a standard deviation of $\sigma_\gamma(0.6) = 1.25$. Imposing these under our discontinuous and joint null give rejection frequencies that are level correct. However, as σ_γ deviates from 1.25 then size distortions are apparent even at the true value of ϕ . For $\sigma_\gamma > 1.25$, the estimated value, lowest p-value, will lead to estimates that are biased downward ($\hat{\phi} < 0.6$). Conversely, $\sigma_\gamma < 1.25$ leads to the estimates that are biased upward ($\hat{\phi} > 0.6$). These results corroborate the findings in Table 2.1, where the bias distortions would lead to incorrect point estimates and poor coverage. The final unobserved value is very important to simulation-based methods, like bootstrap, Monte Carlo, and indirect inference, and may be less relevant for other estimation approaches.

Fig. 2.1.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=100, T=5$

Fig. 2.2.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=100, T=20$

Fig. 2.3.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=300, T=5$

Fig. 2.4.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0$: $N=300, T=20$

Fig. 2.5.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=100, T=5$

Fig. 2.6.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=100, T=20$

Fig. 2.7.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=300, T=5$

Fig. 2.8.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.6$: $N=300, T=20$

Fig. 2.9.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=100, T=5$

Fig. 2.10.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=100, T=20$

Fig. 2.11.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=300, T=5$

Fig. 2.12.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.9$: $N=300, T=20$

Fig. 2.13.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=5$

Fig. 2.14.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=20$

Fig. 2.15.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=5$

Fig. 2.16.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=20$

Fig. 2.17.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=5$

Fig. 2.18.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=100, T=20$

Fig. 2.19.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=5$

Fig. 2.20.: Rejection Frequency for $\alpha = 5\%$ and $\phi = 0.999$: $N=300, T=20$

2.6 Concluding Remarks

Our discontinuous parametric definition of the initial observation leads to provable location-scale invariance of the LSDV estimate, which is suitable for joint consideration of stationary and unit root panels. Our modification results in an indirect infer-

ence objective function that is pivotal, which we invert to construct exact confidence sets. The Monte Carlo method used for our inversion works for our discontinuous joint null, where most, if not all, conventional approaches are restricted to a continuous null for confidence set construction.

In this paper, we focus on the panel-AR(1) model, following the lead of Bun and Kleibergen (2013) and Gouriéroux, Phillips, and Yu (2010), which subsumes that exogenous regressors can be partialled out.

Our simulation study assumed normally distributed errors, but our approach allows for any i.i.d. disturbances that are centered at zero and uncorrelated with the time-invariant errors. Interesting extensions include the assumption of non-Gaussian distributions like the skew-t distribution [Hansen (1994)], the generalized lambda distribution [Gorodnichenko, Mikusheva, and Ng (2012)], or the stable distribution [Beaulieu, Dufour, and Khalaf (2014)]. All of these distributions under specific parameter settings return to a Gaussian distribution as a special case.

3. DYNAMIC PANELS, EXOGENOUS REGRESSORS, AND X-DIFFERENCING

Charles J. Saunders

3.1 Introduction

The incidental parameter bias problem in dynamic panel models is well known. Several methods have been proposed to correct for the bias including, but not limited to, GMM, bias-correction terms, indirect inference, and X-differencing.

Many of the proposed approaches are introduced in the setting of an autoregressive panel without regressors, then assume that standard practices are suitable when extending to the with exogenous regressors case. A similar claim is made regarding the X-differencing method with exogenous regressors in the concluding remarks in Han, Phillips, and Sul (2014).

This claim is examined in the context of the following two dynamic panel models which are distinguished by how the dynamics affect the exogenous regressors. However, both models have a general form used in the derivations that follow. Without loss of generality, this paper focuses on a stationary autoregressive panel with a single exogenous regressor.

Formally, *Model A* is defined by the following system of equations:

$$y_{i,t} = x_{i,t}\beta + \alpha_i + e_{i,t} \tag{3.1}$$

$$e_{i,t} = \phi e_{i,t-1} + \epsilon_{i,t}, \tag{3.2}$$

where $i = 1, \dots, N$, $t = 1, \dots, T$, $\epsilon_{i,t}$ and α_i are i.i.d. and uncorrelated. The exogenous regressors, $x_{i,t}$, are strictly exogenous ($E[\epsilon_{i,t}|x_{i,1}, \dots, x_{i,T}] = 0$ for all t). *Model B* is defined as

$$y_{i,t} = \phi y_{i,t-1} + x_{i,t}\beta + \alpha_i(1 - \phi) + \epsilon_{i,t}, \quad (3.3)$$

where the parameter and series assumptions are identical to the Model A assumptions. The derivations below are based on data with large-N and fixed-T.

3.2 Methods

The X-differencing method constructs a forward-looking version of the model, which is subtracted from the backward-looking model. For consistency, the lagged dependent series must be temporally ahead of the lead series of the forward-looking model.

Combining Equations (3.1) and (3.2), Model A is redefined as:

$$y_{i,t} = \phi y_{i,t-1} + (x_{i,t} - \phi x_{i,t-1})\beta + \alpha_i(1 - \phi) + \epsilon_{i,t}. \quad (3.4)$$

The general form of both models can be described by the following system:

$$y_{i,t} = \phi y_{i,t-1} + z_{i,t}\beta + \alpha_i(1 - \phi) + \epsilon_{i,t} \quad (3.5)$$

$$z_{i,t} = \begin{cases} (x_{i,t} - \phi x_{i,t-1}) & \text{if Model A} \\ x_{i,t} & \text{if Model B.} \end{cases} \quad (3.6)$$

The forward-looking model constructed from Equation 3.5 is

$$y_{i,s} = \phi y_{i,s+1} + z_{i,s}\beta + \alpha_i(1 - \phi) + \epsilon_{i,s}^*, \text{ where} \quad (3.7)$$

$$\epsilon_{i,s}^* = \epsilon_{i,s} + \phi y_{i,s-1} - \phi y_{i,s+1}, \quad (3.8)$$

where $s = 1, \dots, T$. Subtracting Equation 3.8 from Equation 3.5 gives the X-differencing model,

$$(y_{i,t} - y_{i,s}) = \phi(y_{i,t-1} - y_{i,s+1}) + (z_{i,t} - z_{i,s})\beta + (\epsilon_{i,t} - \epsilon_{i,s}^*), \quad (3.9)$$

where $t > s + 1$, and to ensure the model is feasible from the data, $T \geq 5$ for Model A and $T \geq 4$ for Model B. The difference of these two representations eliminates the unobserved heterogeneity (α_i) from the model. In the $\beta = 0$ case, Han et al. (2014) prove that the OLS estimates are unbiased and consistent, since the moment condition $E[(y_{i,t-1} - y_{i,s+1})(\epsilon_{i,t} - \epsilon_{i,s}^*)] = 0$ is true.

When $\beta \neq 0$, the additional moment condition $E[(z_{i,t} - z_{i,s})(\epsilon_{i,t} - \epsilon_{i,s}^*)] = 0$ must also be true for unbiased and consistent parameter estimates. The components that hold under i.i.d. errors or strict exogeneity are: $E[(y_{i,t-1} - y_{i,s+1})\epsilon_{i,t}] = 0$ and $E[(z_{i,t} - z_{i,s})\epsilon_{i,t}] = 0$. Substituting Equation 3.8 into the first moment condition gives:

$$\begin{aligned} -E[(y_{i,t-1} - y_{i,s+1})\epsilon_{i,s}^*] &= -E[(y_{i,t-1} - y_{i,s+1})(\epsilon_{i,s})] \\ &\quad -\phi E[(y_{i,t-1} - y_{i,s+1})(y_{i,s-1} - y_{i,s+1})] \end{aligned} \quad (3.10)$$

Without loss of generality, the uncentered moment of the regressors is defined as $m_2(z) = E[z_{i,t}^2]$ and $m_2(x) = E[x_{i,t}^2]$, the covariances of the regressors are assumed to be zero ($E[x_{i,t}x_{i,r}] = 0$ for all $t \neq r$), stationarity is assumed $|\phi| < 1$ for the dependent series, and the number of time periods is at the allowable minimum the model ($s = t - 3$). These assumptions appear strong, however relaxing these does not change the result.

Assuming that $y_{i,0}$ is the sum of α_i and a second random or nonrandom component, denoted $b_{i,0}$, the recursion of Equation 3.5 gives

$$y_{i,t} = \alpha_i + \phi^t b_{i,0} + \sum_{j=0}^{t-1} \phi^j \epsilon_{i,t-j} + \beta \sum_{j=0}^{t-1} \phi^j z_{i,t-j}. \quad (3.11)$$

From this the first term in Equation 3.10 is

$$-E[(y_{i,t-1} - y_{i,s+1}) \epsilon_{i,s}] = \sigma_\epsilon^2 \phi(1 - \phi). \quad (3.12)$$

The second term in Equation 3.10 is the product of differences between time periods, and these differences are stated below:

$$\begin{aligned} (y_{i,t-1} - y_{i,s+1}) &= \sum_{j=0}^{t-2} \phi^j \epsilon_{i,t-1-j} + \beta \sum_{j=0}^{t-2} \phi^j z_{i,t-1-j} \\ &\quad - \sum_{j=0}^{s-1} \phi^j \epsilon_{i,s+1-j} - \beta \sum_{j=0}^{s-1} \phi^j z_{i,s+1-j} \end{aligned} \quad (3.13)$$

$$\begin{aligned} (y_{i,s-1} - y_{i,s+1}) &= \sum_{j=0}^{s-2} \phi^j \epsilon_{i,s-1-j} + \beta \sum_{j=0}^{s-2} \phi^j z_{i,s-1-j} \\ &\quad - \sum_{j=0}^{s-1} \phi^j \epsilon_{i,s+1-j} - \beta \sum_{j=0}^{s-1} \phi^j z_{i,s+1-j}. \end{aligned} \quad (3.14)$$

The assumption of strict exogeneity allows for the moment to be examined as independent parts, where

$$-\phi E \left[\left(\sum_{j=0}^{t-2} \phi^j \epsilon_{i,t-1-j} - \sum_{j=0}^{s-1} \phi^j \epsilon_{i,s+1-j} \right) \left(\sum_{j=0}^{s-2} \phi^j \epsilon_{i,s-1-j} - \sum_{j=0}^{s-1} \phi^j \epsilon_{i,s+1-j} \right) \right]$$

equals $\sigma_\epsilon^2 \phi(\phi - 1)$, which is the result from Han et al. (2014) and added to Equation 3.12 equals zero.¹ So the inclusion of exogenous regressors does not affect the moments derived from the idiosyncratic errors.

The moment condition derived from the product of Equations 3.13 and 3.13 for the exogenous regressors gives

$$-\phi \beta^2 E \left[\left(\sum_{j=0}^{t-2} \phi^j z_{i,t-1-j} - \sum_{j=0}^{s-1} \phi^j z_{i,s+1-j} \right) \left(\sum_{j=0}^{s-2} \phi^j z_{i,s-1-j} - \sum_{j=0}^{s-1} \phi^j z_{i,s+1-j} \right) \right]$$

---

¹I assume that t is large enough so that $\frac{1-\phi^{2t}}{1-\phi^2} \approx \frac{1}{1-\phi^2}$, exact finite sample derivations would need to relax this assumption.

and under the assumptions above equals $m_2(z)\beta^2\phi(\phi-1)$ for model B. Collecting the moments for model B gives,

$$-E[(y_{i,t-1} - y_{i,s+1})\epsilon_{i,s}^*] = m_2(z)\beta^2\phi(\phi-1). \quad (3.15)$$

Substituting equation 3.8 into the second moment condition and simplifying gives

$$E[(z_{i,t} - z_{i,s})\epsilon_{i,s}^*] = E[(z_{i,t} - z_{i,s})(\epsilon_{i,s} + \phi y_{i,s-1} - \phi y_{i,s+1})] \quad (3.16)$$

$$= \phi^2\beta m_2(z). \quad (3.17)$$

For model B where $m_2(x) = m_2(z)$, the presence of exogenous regressors ($\beta \neq 0$) violates both moment conditions in estimation of ϕ and β from equation 3.9.

Model A moment conditions are affected by the lagged exogenous regressors in the definition of $z_{i,t}$, derived as

$$-E[(y_{i,t-1} - y_{i,s+1})\epsilon_{i,s}^*] = m_2(x)\beta^2\phi \quad (3.18)$$

$$E[(z_{i,t} - z_{i,s})\epsilon_{i,s}^*] = \phi^2\beta m_2(x). \quad (3.19)$$

3.3 Simulations

To illustrate the bias of X-differencing estimation in the presence of exogenous regressors, the most basic simulation study is employed. The data generating process is restricted to Model B, specifically,

$$y_{i,t} = \phi y_{i,t-1} + x_{i,t}\beta + \alpha_i(1 - \phi) + \epsilon_{i,t} \text{ and} \quad (3.20)$$

$$x_{i,t} = \nu_{i,t}, \quad (3.21)$$

where α , ϵ and ν are drawn from an i.i.d. standard normal distribution, the number of time periods is fixed ($T = 4$), $\phi \in \{0.4, 0.8\}$, and $\beta \in \{0.5, 1\}$. The X-differencing representation is defined by Equation 3.9.

The first estimation method is OLS, which is comparable with the approach outlined above. The second estimation method is a feasible generalized least squares (FGLS) approach, which requires consistent initial parameter estimates. Han et al. (2014) contend that OLS under the assumption of no dynamics ($\phi = 0$) is suitable when implemented in the FGLS approach of Bhargava et al. (1982). Specifically, the estimated errors then form the basis of the X-differencing estimate of ϕ . The equations of interest for FGLS are

$$\tilde{\eta}_{i,t} = y_{i,t} - x_{i,t}\tilde{\beta} - \tilde{\alpha}_i(1 - \phi) \text{ and} \quad (3.22)$$

$$(\tilde{\eta}_{i,t} - \tilde{\eta}_{i,s}) = \phi(\tilde{\eta}_{i,t-1} - \tilde{\eta}_{i,s+1}) + (\epsilon_{i,t} - \epsilon_{i,s}^*), \quad (3.23)$$

where the ‘tilde’ items are obtained from the initial static panel OLS estimates. The resulting estimate of ϕ is used to construct a weighting matrix, Ω , with each element defined by $\omega_{k,j} = (\phi^{|k-j|})/(1 - \phi^2)$. The estimate of Ω from each stage is then used in later stages until convergence is achieved. The simulations presented focus on the initial OLS estimate (denoted below as GLS1), since this initial consistency is imperative for an appropriate FGLS estimate.

Table 3.1: X-differencing Parameter Estimation Bias

N	ϕ			β		
	True	OLS	GLS1	True	OLS	GLS1
200	0.4	-0.0465	0.0063	0.5	-0.0355	-0.0779
200	0.4	-0.1061	0.0318	1.0	-0.0721	-0.1552
200	0.8	-0.0634	0.0481	0.5	-0.1533	-0.1904
200	0.8	-0.1518	0.1680	1.0	-0.3127	-0.3813
300	0.4	-0.0360	0.0080	0.5	-0.0285	-0.0707
300	0.4	-0.0919	0.0314	1.0	-0.0544	-0.1418
300	0.8	-0.0387	0.0512	0.5	-0.1307	-0.1757
300	0.8	-0.0997	0.1713	1.0	-0.2601	-0.3513

Simulations based on 1,000 replications.

The OLS estimation of Equation 3.20 results in parameter estimates that are downward biased, which matches the expectation from the derived moment condi-

tions. The OLS bias increases in magnitude as β increases. The OLS bias increases as ϕ increases, but the $\phi(1 - \phi)$ portion of the derived bias indicates that this should decrease as $\phi \rightarrow 1$.

The initial GLS estimates for β are biased downward, which is a consequence of imposing the LSDV (or fixed effects) estimator on a stationary panel with autoregressive errors. The demeaned autoregressive errors are correlated with the demeaned exogenous regressors, specifically $E[\eta_{i,t}^* x_{i,t}^*] \neq 0$. This bias affects the initial estimate of ϕ via X-differencing, resulting in an upward bias.

The magnitude of the bias for both estimators depends heavily on the parameters of the exogenous regressors, where higher values increase the magnitude.

3.4 Final Remarks

This paper demonstrates that the inclusion of exogenous regressors into the X-differencing framework can be problematic. OLS applied to the X-differencing model with exogenous regressors leads to downward biased parameter estimates. The initial estimate for the GLS approach provides some improvement for the lagged dependent parameter, but the parameters for the exogenous regressors are more downward biased than OLS.

This note focuses on a stationary autoregressive panel with exogenous regressors, but the X-differencing method has been shown to be much more versatile. Future consideration of the effect of exogenous regressors in the X-differencing framework should take into account unit root and higher-order dynamic panels, exogenous regressor covariances, and the number of time periods.

4. ON AUTOREGRESSIVE PANELS WITH EXOGENOUS REGRESSORS: INVARIANCE, INDIRECT INFERENCE, AND CONFIDENCE SETS

Charles J. Saunders

4.1 Introduction

Individual heterogeneity in panel data leads to the incidental parameter bias as dynamics is introduced, first identified in Neyman and Scott (1948). For a fixed number of time periods, the magnitude of the bias can be significant and is not eliminated as the number of cross-sections increase [Nickell (1981)]. The majority of papers focus solely on the incidental parameter bias, while the dynamic panel model with exogenous regressors has mainly been addressed by cursory discussion with respect to conventional approaches drawn from the time series literature. Approaches that exploit panel data structures to account for exogenous regressors are relatively modest.

Covariates are addressed in the instrumental variable (IV) and generalized method of moment (GMM) literature under assumptions of strong exogeneity or predetermined exogenous regressors. The most popular estimators are Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), and Blundell and Bond (1998). These methods include the exogenous variables in the set of valid instruments. The GMM estimators generally rely on Large- N and Small- T asymptotics for both consistency and efficiency.

In the context of dynamic panels, Phillips and Sul (2007) showed that the least squares dummy variable (LSDV) estimator remains biased when relevant covariates are included in the model. The inclusion of covariates has a bias-reducing effect, which

depends on the magnitude of the absolute value of the covariate parameters, and the *variance-ratio* of the covariates and the idiosyncratic errors. The variance-ratio is a nuisance parameter on the LSDV estimates of the lagged dependent parameter and interferes inversely with the estimate of the covariate parameters. A potential solution fixes the variance-ratio by assumption, but the true variance-ratio is generally unobserved and difficult to estimate consistently.

In the indirect inference context, Gouriéroux, Phillips, and Yu (2010, pg. 74) construct a projection matrix to partial out the exogenous regressors. The model derived with the projection matrix is used in their simulation stage of the indirect inference estimator. Their projection method conforms with the Frisch-Waugh-Lovell (FWL) theorem [Frisch and Waugh (1933) and Lovell (1963)], so the estimate of the lagged dependent parameter is unchanged. The associated bias is unaffected by this partialling out, so the covariate parameters and variance ratio remain as nuisance parameters in the estimate. As shown in Khalaf and Saunders (2014), the indirect inference estimator applied to dynamic panels must have explicit assumptions on the initial observation, so the simulations associated with indirect inference require assumptions on the joint distribution of the initial dependent observation and exogenous regressors.

The first contribution of this chapter is the introduction of the augmented least-squares dummy variable estimator (augmented-LSDV) to dynamic panel models with exogenous regressors (panel-ARX(1,K)). Augmenting the LSDV estimator in this manner still partials out the exogenous regressors in the usual sense, but goes further to eliminate the covariate parameters from the bias. The resulting augmented-LSDV estimate has an equivalent distribution as the LSDV estimator applied to a panel-AR(1) model, which allows the augmented-LSDV estimator to inherit simplified invariance and nuisance parameter properties.

The augmentation of the proposed estimator is a weighting matrix that is a block diagonal matrix for each cross-section, where the block consists of all the exogenous regressor observations for the cross-section. The use of a block diagonal matrix is com-

mon to several GMM/IV dynamic panel methods in the construction of the instrument matrix. The blocks in these approaches are commonly lags or contemporaneous observations of the dependent variable, with some inclusion of lead observations for strongly exogenous regressors. The proposed matrix includes all exogenous regressors, which is not without precedent. The Chamberlain device in Griliches and Intriligator (1984, Chapter 22) defines a linear relationship between the time-invariant coefficient and the time-invariant exogenous regressors for all time periods. The Chamberlain device, however, is not block diagonal. The inclusion of both lags and lead into the estimation has ties to the Dynamic OLS of Stock and Watson (1993). The proposed augmented-LSDV estimator allows for the inclusion of all lags and leads, so long as the data consists of a large number of cross-sections relative to regressors and time periods.

The second contribution is the extension of the indirect inference estimator (IIE) of Gouriéroux, Phillips, and Yu (2010) with the proposed augmented-LSDV to achieve consistent parameter estimates in the presence of exogenous regressors. The paper extends the initial observation discontinuous function approach, put forth in Khalaf and Saunders (2014), to account for exogenous regressors and ensures that the augmented-IIE method is nuisance parameter free. The augmented-LSDV estimator focuses the estimation on the lagged dependent parameter where the invariance theorems lead to simplified simulation assumptions specifically a panel-AR(1) model is injective at the true lagged dependent parameter value. A second stage LSDV estimation conditional on the augmented-IIE point estimate allows for the recovery of the exogenous variable parameters. A small Monte Carlo study is used to examine the bias of several approaches, including system-GMM [Blundell and Bond (1998)], LSDV, the approach proposed in Gouriéroux, Phillips, and Yu (2010), two variants of the indirect inference estimator, and the proposed augmented-IIE. The augmented-IIE, proposed in this paper, provides the best bias-correction for the parameters of the model.

The third contribution extends the indirect confidence set inference method of Khalaf and Saunders (2014) to allow for exogenous regressors; the augmented-LSDV estimator is used to obtain the estimate from the data. The method constructs exact confidence sets for the parameters by inverting the indirect inference objective function using a second round of Monte Carlo replications to standardize the distribution of the objective function. Level control is ensured since the IIE and Monte Carlo simulations generated are exchangeable under the null. The confidence set for exogenous regressor parameters is collected by a second stage LSDV estimate conditional on a given lagged dependent parameter value. The joint confidence set of the parameters is constructed as the infimum and supremum of the exogenous regressor parameters over the confidence set for the lagged dependent parameter, as outlined in Dufour (1990).

Lastly, the indirect confidence set inference properties are examined in the context of a Monte Carlo simulation study. The study is constructed to incorporate an exogenous regressor¹ and allowing for varying degrees of persistence: no dynamics, stationary, local-to-unity, and the unit root. The rejection frequencies of the joint confidence set are level correct under the null with relatively few cross-sections and time periods. Under the alternative, the rejection frequencies increase with the number of cross-sections and time periods. The proposed method exhibits robust rejection properties under alternative exogenous regressor dynamics and model specifications including a collapsed version of the augmented-LSDV estimator. The coverage of the exogenous parameter confidence sets are also examined.

The chapter is organized as follows, the panel-ARX(1,K) model is formally introduced in Section 4.2. Section 4.3 introduces the augmented-LSDV estimator, and the construction of the projection matrix from the exogenous regressors. The invariance properties of the augmented-LSDV estimator are defined in Section 4.4. The augmented-LSDV is used to formally extend the indirect inference estimator, in Section 4.5, to ensure unbiased estimates, which is demonstrated in a small Monte

---

¹Multiple regressors is covered by the theory presented below.

Carlo study. In Section 4.6 extends the indirect confidence set inference method of Khalaf and Saunders (2014) to allow for exogenous regressors, and demonstrate, in Section 4.7, the rejection frequency properties through a simulation study. Section 4.8 concludes the chapter and highlights some worthwhile extensions of the proposed method.

4.2 Framework

The general form of a dynamic panel model with exogenous regressors, denoted as panel-ARX(1,K), is defined by,

$$\begin{aligned} y_{i,t} &= (1 - \phi) (\mu_i + x'_{i,0}\beta) + \phi y_{i,t-1} + x'_{i,t}\beta + \epsilon_{i,t} \\ i &= 1, \dots, N \text{ and } t = 1, \dots, T, \end{aligned} \quad (4.1)$$

where i indicates the cross-section of the panel, t is the time index of the observations, $y_{i,t}$ is the dependent variable, and $x_{i,t}$ is a vector of K exogenous regressors (indexed by k where $k = 1, \dots, K$, leading to $x_{i,t,k}$), and β is a $K \times 1$ vector of parameters. The initial period exogenous realizations, $x_{i,0}$, are unobserved since the time index is less than one.

The time-invariant error terms, μ_i , can be either random or non-random. The time-varying errors terms, $\epsilon_{i,t}$, must follow a distribution that allows for independent draws, and is centered at zero. The error terms are assumed to be contemporaneously uncorrelated,

$$E[\epsilon_{i,t} | \mu_j] = 0 \quad \forall i, j.$$

It is common to assume both error terms follow independent Gaussian distributions, but the framework is not restricted to the Gaussian family.

The exogenous regressors in this model are assumed to be strongly exogenous, specifically

$$E[\epsilon_{j,s}|x_{i,t,k}] = 0 \quad \forall t, s \text{ and } i, j.$$

This assumption can be relaxed to allow for certain forms of endogeneity, for example correlation between the regressor and the unobserved heterogeneity. The LSDV estimator allows for correlation between $x_{i,t,k}$ and μ_j for all $i, j = 1, \dots, N$, where alternative methods of estimation are less robust to this correlation.

Without loss of generality, the analysis below focuses on a panel-ARX(1,1) model,

$$y_{i,t} = (1 - \phi) (\mu_i + x_{i,0}\beta) + \phi y_{i,t-1} + x_{i,t}\beta + \epsilon_{i,t}, \quad (4.2)$$

with a single exogenous regressor ($K = 1$). Increasing the number of exogenous regressors affects the complexity of the derivations, but the structure of the proposed method is retained.

4.2.1 LSDV with Exogenous Regressors

The general form of the LSDV estimator demeans all the series to remove the time-invariant coefficient,

$$(y_{i,t} - y_{i,\bullet}) = \phi(y_{i,t-1} - y_{i,\bullet-1}) + (x_{i,t} - x_{i,\bullet})\beta + (\epsilon_{i,t} - \epsilon_{i,\bullet})$$

where the sample means are given by

$$\begin{aligned} y_{i,\bullet} &= \frac{1}{T-1} \sum_{t=2}^T y_{i,t}, & x_{i,\bullet} &= \frac{1}{T-1} \sum_{t=2}^T x_{i,t}, \\ y_{i,\bullet-1} &= \frac{1}{T-1} \sum_{t=1}^{T-1} y_{i,t} \quad \text{and} \quad \epsilon_{i,\bullet} &= \frac{1}{T-1} \sum_{t=2}^T \epsilon_{i,t}. \end{aligned}$$

To simplify the notation, and through appropriate stacking of observations to obtain the vector notation representation, the demeaned series will be denoted with an asterisk (*). The estimated model can be represented as

$$y^* = y_{-1}^* \phi + x^* \beta + \epsilon^*, \quad (4.3)$$

and the “−1” indicates the degree of lag in the time index. The LSDV estimates of the parameters from this model are computed as

$$\begin{aligned} \begin{pmatrix} \ddot{\phi} \\ \ddot{\beta} \end{pmatrix} &= \begin{bmatrix} y_{-1}^{*'} y_{-1}^* & y_{-1}^{*'} x^* \\ x^{*'} y_{-1}^* & x^{*'} x^* \end{bmatrix}^{-1} \begin{pmatrix} y_{-1}^{*'} y^* \\ x^{*'} y^* \end{pmatrix} \\ \begin{pmatrix} \ddot{\phi} \\ \ddot{\beta} \end{pmatrix} &= \begin{pmatrix} \phi \\ \beta \end{pmatrix} + \begin{bmatrix} y_{-1}^{*'} y_{-1}^* & y_{-1}^{*'} x^* \\ x^{*'} y_{-1}^* & x^{*'} x^* \end{bmatrix}^{-1} \begin{pmatrix} y_{-1}^{*'} \epsilon^* \\ x^{*'} \epsilon^* \end{pmatrix}. \end{aligned} \quad (4.4)$$

Consistent estimates of ϕ and β require that certain moment conditions are zero in expectation. Under the assumption of strict exogeneity, the following moment conditions hold: $E[x^{*'} \epsilon^*] = 0$ and $E[y_{-1}^{*'} \epsilon^*] = 0$. However, $E[y_{-1}^{*'} \epsilon^*] \neq 0$ due to three non-zero moment conditions:

$$E_i[y_{i,t-1} \epsilon_{i,t}] = 0, \quad E_i[y_{i,t-1} \epsilon_{i,\bullet}] \neq 0, \quad E_i[y_{i,\bullet-1} \epsilon_{i,t}] \neq 0, \quad \text{and} \quad E_i[y_{i,\bullet-1} \epsilon_{i,\bullet}] \neq 0.$$

The violation of the latter three moment conditions leads to biased LSDV estimates, when T is fixed. Nickell (1981) derived the analytical form of this bias without exogenous regressors. Phillips and Sul (2007) demonstrate that the presence of exogenous regressors, such that $\beta \neq 0$, reduces magnitude of the bias on ϕ . The bias is not only affected by β , but also the variance ratio of x and ϵ , denoted as $\sigma_x^2/\sigma_\epsilon^2$, where both conditions are derived from the covariance of y_{-1}^* and x^* .

A conventional approach is to partial out the exogenous regressors, which reduces the confounding effect of exogenous regressors on the bias. Let M_x^* be a projection matrix defined as

$$M_x^* = I_{N(T-1)} - x^*(x^{*'}x^*)^{-1}x^{*'},$$

which multiplied to equation 4.3 partials out x^* ,

$$\begin{aligned} M_x^*y^* &= M_x^*y_{-1}^*\phi + M_x^*x^*\beta + M_x^*\epsilon^* \\ M_x^*y^* &= M_x^*y_{-1}^*\phi + M_x^*\epsilon^*. \end{aligned} \tag{4.5}$$

This is the conventional approach, however recursive substitution shows that this has not removed β from the equation,

$$M_x^*y^* = M_x^*y_{-2}^*\phi^2 + M_x^*x_{-1}^*\beta\phi + M_x^*\epsilon_{-1}^*\phi + M_x^*\epsilon^*.$$

The matrix M_x^* is defined using $x_{i,2}^*, \dots, x_{i,T}^*$, but x_{-1}^* is a vector that includes $x_{i,1}^*$ leading to,

$$M_x^*x_{-1}^* \neq 0.$$

So in the LSDV framework, the value of β influences the estimation of the other parameters, and this effect is not removed by conventional partialling out approaches. By the FWL theorem, the partialling out method does not change the estimate of ϕ , which indicates that the bias-reducing effect of exogenous regressors is not removed with conventional methods.

Gouriéroux, Phillips, and Yu (2010, pg. 74) indicate that partialling out the exogenous regressors leads to the application of the IIE method using Equation 4.5. Their method of partialling out is as follows,

$$\begin{aligned} X_i' &= [x_{i,2}, \dots, x_{i,T}] \\ Z_i &= [\iota_T, X_i] \\ P_i &= I_T - Z_i(Z_i'Z_i)^{-1}Z_i' \\ P_i y_i &= P_i y_{i,-1}\phi + P_i \epsilon_i. \end{aligned}$$

The intuition of their approach is to regress each cross-section on a constant and the exogenous regressors for the cross-section. As above, through recursive substitution this method does not take into account the $x_{i,1}$ observation,

$$P_i y_i = P_i y_{i,-2}\phi^2 + P_i x_{i,-1}\beta\phi + P_i \epsilon_i\phi + P_i \epsilon_i,$$

and consequently $P_i x_{i,-1} \neq 0$. Since β remains in the equation, then it will still influence the estimate of ϕ .

The indirect inference simulations of a dynamic panel model with exogenous regressors requires consistent estimates of β and σ_ϵ^2 . Due to the inherent bias of the model, the LSDV estimator given by Equation 4.4 does not consistently estimate these parameters even when the conventional partialling out method and the partialling out approach of Gouriéroux, Phillips, and Yu (2010) are employed.

4.2.2 A Comment on Kiviet and van den Doel (1994) Invariance

The following model,

$$\begin{aligned} y_{i,1} &= \mu_i + x_{i,1}\beta + \frac{1}{\sqrt{1-\phi^2}}\epsilon_{i,1} \\ y_{i,t} &= \phi y_{i,t-j} + (1-\phi)\mu_i + x'_{i,t}\beta(1-\phi) + \epsilon_{i,t}, \end{aligned}$$

is presented in Kiviet and van den Doel (1994), which considers only the observed data. The authors present two null hypotheses and claim that their alternative regression model leads to invariance to y_1 , β , μ , and σ_ϵ^2 . The first null is defined as $H_{0,1} : \phi = 0$, and the second null is defined as $H_{0,2} : \phi = \phi_0$ where $\phi_0 \in (-1, 1)$.

The altered regression model of Kiviet and van den Doel (1994) includes the lags of the exogenous regressors and cross-section effects. In their paper, the authors stack the time periods first then the cross-sections, which is the opposite order to the rest of this thesis but this is a cosmetic distinction. To remain close to their presentation, only this section presents the model in their alternative matrix order. A $T \times T$ lag operator matrix is defined as

$$L = \begin{bmatrix} 0' & 0 \\ I_{T-1} & 0 \end{bmatrix}.$$

The matrix representation of the alternative model is,

$$\begin{aligned} y_1 &= \mu + x_1\beta + \frac{1}{\sqrt{1-\phi^2}}\epsilon_1 \\ y &= \phi y_{-1} + (\iota_T \otimes \mu)(1-\phi) + x\beta(1-\phi) \\ &\quad + (L \otimes I_N)(\iota_T \otimes \mu)(1-\phi) + (L \otimes I_N)x\beta(1-\phi) + \epsilon, \end{aligned}$$

which is used to build the projection matrix. Kiviet and van den Doel (1994) also define y_{-1} from the recursion of the series as,

$$\begin{aligned} y_{-1} &= (\Gamma_1(\phi) \otimes I_N)y_1 + (\Gamma(\phi) \otimes I_N)(\iota_T \otimes \mu)(1-\phi) \\ &\quad + (\Gamma(\phi) \otimes I_N)x\beta(1-\phi) + (\Gamma(\phi) \otimes I_N)\epsilon \end{aligned}$$

where

$$\Gamma_1(\phi) = \begin{bmatrix} 1 \\ \phi \\ \vdots \\ \phi^{T-1} \end{bmatrix} \quad \text{and} \quad \Gamma(\phi) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \phi & 1 & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \phi^{T-1} & \cdots & \phi & 1 \end{bmatrix}$$

and under $H_{0,1}$ these simplify to $\Gamma(0) = I_T$ and $\Gamma_1(0) = (1 - L)\iota_T$.

Let V be defined as:

$$V = [x : (L\Gamma(\phi) \otimes I_N)x : (\iota_T \otimes I_N) : (L\Gamma(\phi)\iota_T \otimes I_N)],$$

which is used to build the projection matrix, $M_V = I_{NT} - V(V'V)^{-1}V'$. Clearly, this projection matrix will partial out the exogenous regressors and fixed effects from the equation above, leaving $M_V y_{-1}$ for careful consideration. The exogenous and fixed effects encapsulated in y_{-1} are also partialled out by the M_V matrix, which leaves the initial observation.

Kiviet and van den Doel (1994) show that under $H_{0,1} : \phi = 0$, the modified regression equation is invariant to y_1 . Under $H_{0,2} : \phi = \phi_0$ this requires the initial values to be either net zero over all time periods, or time period independent. The influence of the initial values on the series is given by, $(\iota_T \otimes I_N) - (L\Gamma(\phi)\iota_T \otimes I_N) = (1 - L\Gamma(\phi))\iota_T \otimes I_N$. For invariance it must be the case that $(1 - L\Gamma(\phi))\iota_T = (1 - L)\iota_T$. Kiviet and van den Doel (1994) show that this is true under $H_{0,1}$ or $\phi = 0$. However, when $\phi \neq 0$ then $\Gamma(\phi)$ does not equal the identity matrix so this transformation is not invariant to y_1 under $H_{0,2}$. As a result, under $H_{0,2}$ the estimation and inference for ϕ will be affected by magnitude of $x_{i,0}$, β , and assumptions on σ_γ .

4.3 Augmented-LSDV

To account for the $x_{i,1}$ observation, I introduce the augmented-LSDV estimator. The properties of this new estimator are presented in the following sections. The

augmented-LSDV estimator exploits panel structure, specifically the number of cross-sections, to allow for complete partialling out of β .

The importance of the initial observation assumptions is documented in Khalaf and Saunders (2014). This paper extends their work to include exogenous regressors with the following assumption, which includes the initial period exogenous regressors and associated parameter (β), the long-run variance of the series (σ_γ), and the disturbances.

Assumption 4 (Initial Values for Simulated Series) *The initial value of the series for any i is defined by*

$$y_{i,0} = \mu_i + x_{i,0}\beta + \sigma_\gamma\epsilon_{i,0}$$

where $\sigma_\gamma \geq 0$, and $\epsilon_{i,0}$ follows the same distribution as $\epsilon_{i,t}$ and $E[\epsilon_{i,0}\epsilon_{i,t}] = 0$ for all t , and $x_{i,0}$ is strongly exogenous.

The proofs of the theorems and corollaries are based on the recursive decomposition of the demeaned series, so stated here as Lemma 2.

Lemma 2 (Recursive Decomposition) *Under assumption 4, the demeaned series by recursive substitution is*

$$y_{i,t}^* = a_{i,t}(\phi, x_i, t, T)\beta + g_{i,t}(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T), \text{ where} \quad (4.6)$$

$$a_{i,t}(\phi, x_i, t, T) = \left(\sum_{j=0}^{t-1} x'_{i,t-j}\phi^j - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} x'_{i,s-j}\phi^j \right) \text{ and} \quad (4.7)$$

$$\begin{aligned} g_{i,t}(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) &= \left(\sum_{j=0}^{t-1} \epsilon_{i,t-j}\phi^j - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} \epsilon_{i,s-j}\phi^j \right) \\ &\quad + \left(\sigma_\gamma\epsilon_{i,0}\phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma\epsilon_{i,0}\phi^s \right). \end{aligned} \quad (4.8)$$

Let $g_{i,t} = g_{i,t}(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T)$, $g_{i,t-1} = g_{i,t-1}(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t-1, T)$, $a_{i,t} = a_{i,t}(\phi, x_i, t, T)$, and $a_{i,t-1} = a_{i,t-1}(\phi, x_i, t-1, T)$, where the vector representations are g , g_{-1} , A , and A_{-1} , respectively. The vector representation of the demeaned model is,

$$y^* = A\beta + g. \quad (4.9)$$

Under assumption 4 and by Lemma 2, the demeaned series consists of a term ($A\beta$) that is a function of β , ϕ and the exogenous regressors, and a term (g) that is a function of ϕ , σ_γ , and the disturbances (ϵ). The sum operator is linear, so Equation (4.7) can be redefined as a weighted sum of $x'_{i,1}, \dots, x'_{i,T}$, where the weight on an individual $x_{i,t}$ are nonlinear and are well-defined functions of ϕ , t , and T .²

Let $\omega = (\omega'_2, \omega'_3, \dots, \omega'_T)'$ be the common weight vector for all cross-sections where $\omega_t = (\omega_{t,1}, \omega_{t,2}, \dots, \omega_{t,T})'$, and each element in the vector is defined as

$$\omega_{t,r} \begin{cases} = \phi \left(\frac{1-\phi^{T+1-r}}{(T-1)(1-\phi)} + \phi^{t-r-1} \right) & \text{if } r = 1 \\ = \left(\frac{1-\phi^{T+1-r}}{(T-1)(1-\phi)} + \phi^{t-r} \right) & \text{if } 2 \leq r \leq t \\ = \left(\frac{1-\phi^{T+1-r}}{(T-1)(1-\phi)} \right) & \text{if } r > t, \end{cases} \quad (4.10)$$

for $r = 1, \dots, T$.

Let $X_{i,k} = (x_{i,1,k}, \dots, x_{i,T,k})'$ and $X_i = (X'_{i,1}, \dots, X'_{i,K})'$, so for a given cross-section, i , I can define the following expression,

$$A = \begin{bmatrix} (I_{(T-1)} \otimes X'_1) \\ (I_{(T-1)} \otimes X'_2) \\ \vdots \\ (I_{(T-1)} \otimes X'_N) \end{bmatrix} [\iota_K \otimes \omega] = W_X W_\omega.$$

---

²A similar redefinition of equation (4.8) is possible in terms of ϵ , but this is not required for proofs or derivation of the augmented-LSDV estimator.

It is important to note that the vector ω is identical for all the cross-sections and each exogenous regressor. Leading to an alternative representation of Equation (4.9),

$$y^* = W_X W_\omega \beta + g. \quad (4.11)$$

The demeaned lagged series follows a similar derivation and is provided in the Appendix.

Assuming that W_X is full column rank, $rank(W_X) = KT(T - 1)$, then it is possible to partial out the exogenous regressors. The following section formalizes this assumption and implications. The matrix M_X is defined as

$$M_X = I_{N(T-1)} - W_X(W_X' W_X)^{-1} W_X', \quad (4.12)$$

which leads to the following set of identities:

$$M_X x^* = 0, \quad M_X A = 0, \quad \text{and} \quad M_X A_{-1} = 0.$$

These identities are used to derive the augmented-LSDV estimator, beginning with demeaned representation given by equation (4.3),

$$\begin{aligned} y^* &= y_{-1}^* \phi + x^* \beta + \epsilon^* \\ M_X y^* &= M_X y_{-1}^* \phi + M_X x^* \beta + M_X \epsilon^* \\ M_X y^* &= M_X y_{-1}^* \phi + M_X \epsilon^* \\ M_X (W_X W_\omega \beta + g) &= M_X (W_X W_{\omega, -1} \beta + g_{-1}) \phi + M_X \epsilon^* \\ M_X g &= M_X g_{-1} \phi + M_X \epsilon^*. \end{aligned}$$

Although this final representation cannot be constructed from the data, the representations lead to the following representations of the augmented-LSDV estimates of ϕ :

$$\begin{aligned}\hat{\phi} &= (y_{-1}^{*'} M_X y_{-1}^*)^{-1} y_{-1}^{*'} M_X y^* \\ &= (g_{-1}' M_X g_{-1})^{-1} g_{-1}' M_X g \\ &= \phi + (g_{-1}' M_X g_{-1})^{-1} g_{-1}' M_X \epsilon^*.\end{aligned}$$

All these estimates are numerically equal, and the first can be constructed from the data. The last representation demonstrates that the augmented-LSDV estimator is not a function of β , since β is not an argument in g , g_{-1} , or M_X .

Based on the structure of the projection matrix, the augmented-LSDV estimator could be compared to the Chamberlain device. However, the objectives of these methods are quite different. The Chamberlain device is the regression of the estimated unobserved heterogeneity on all the exogenous regressors for all time periods, to account for endogeneity between the unobserved heterogeneity and the exogenous regressors. The augmented-LSDV estimator is used to account for the correlation between the demeaned dependent series and the exogenous regressors.

The W_X matrix is constructed such that for each time period the demeaned dependent series is regressed on every exogenous regressor observation by cross-section. The augmented-LSDV estimator can be represented in two stages. In the first stage, for a given cross-section, the augmented-LSDV estimator imposes the following model,

$$y_{i,t}^* = x'_{i,1} \delta_{t,1} + x'_{i,2} \delta_{t,2} + \dots + x'_{i,T} \delta_{t,T} + u_{i,t}, \quad (4.13)$$

so the coefficients (denoted $\delta_{t,r}$) for all time periods are indexed by t twice, which results in $2TK$ parameters. The coefficients in this model are $\delta_{t,r} = \omega_{t,r} \beta$. The second stage regresses the errors on their lags, specifically

$$u_{i,t} = \phi u_{i,t-1} + \varepsilon_{i,t}^*, \quad (4.14)$$

where the vector of errors from this model are $\varepsilon^* = M_X \epsilon^*$. The augmented-LSDV estimation replaces $u_{i,t}$ with their estimated counterparts to estimate ϕ .

4.3.1 Projection Matrix Properties

The augmented-LSDV estimator relies on the projection matrix, equation (4.12), so the properties of the W_X matrix must be defined. As shown above, the W_X matrix is derived from Equation 4.7 resulting in a block diagonal structure. The W_X matrix exploits the large number of cross-sections relative to the number of time periods feature of dynamic panel models.

Assumption 5 (Rank Condition for Exogenous Projection Matrix) *The number of cross-sections are sufficiently large to ensure that the matrix,*

$$W_X = \begin{bmatrix} (I_{(T-1)} \otimes X'_1) \\ (I_{(T-1)} \otimes X'_2) \\ \vdots \\ (I_{(T-1)} \otimes X'_N) \end{bmatrix} \quad (4.15)$$

is full column rank, $\text{rank}(W_X) = KT(T-1)$. The minimum number of cross-sections restriction is

$$N > KT$$

where T is the number of time periods and K is the number of exogenous regressors not including cross-section fixed effects.

Assumption 5 ensures that the product, $W_X'W_X$, is invertible. Assume that a square, full-rank permutation matrix (D) can be constructed such that,

$$DW_X = I_{(T-1)} \otimes X, \quad \text{where} \quad X = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_N' \end{bmatrix}.$$

The identity matrix is full rank, so I need only focus on the rank condition of X , which is an $N \times KT$ matrix of exogenous regressors. The full column rank condition of assumption 5 simplifies to $\text{rank}(X) = KT < N$.

Satisfying assumption 5 puts restrictions on the type of exogenous regressors that can be included. Exogenous variables that vary sufficiently in both i and t are suitable for inclusion in X . Exogenous variables that vary only in t (e.g. common regressor) will generate columns in X that are collinear, so all but one column would be excluded from X . This column will have the same function as a constant term, which has already been removed by demeaning the series. Exogenous variables that vary only in i also generate collinear columns in X . Since the estimator demeans them these variables will be part of the LSDV estimator so excluded entirely. Dummy variables would be special cases and the rank condition would have to hold for inclusion.

The augmented-LSDV uses a block diagonal structure to capture all of the correlation of the exogenous regressors with the dependent and lagged dependent variables. The method exploits the double index of the panel-ARX model and specifically the large- N and fixed- T with respect to assumption 5. Unique to W_X is the inclusion of all past, present and future exogenous regressor observations that define the blocks.

The augmented-LSDV estimator is appropriate for data where the number of cross-sections greatly exceeds the number of time periods times exogenous regressors. The augmented-LSDV estimator may not be suitable for smaller panels, since these may violate assumption 5. The proposed estimator is also unique to dynamic panels, since the projection matrix for time series and vector autoregression models cannot

be constructed and satisfy assumption 5. So these methods must rely on a large number of time period asymptotics. For these models, the complete partialling out of the parameters on the exogenous regressors is not possible.

Careful examination of Equation (4.10) (and (B.7) in the Appendix) shows that some of the nonlinear weights are common for different time periods. Consider the case where $t < s < r$ then

$$\omega_{t,r} = \left(\frac{1 - \phi^{T+1-r}}{(T-1)(1-\phi)} \right) = \omega_{s,r},$$

so both t and s have common weights. With the collection of these lead exogenous regressors to the same column and accounting for the lagged dependent common weights, a $W_X^{collapsed}$ matrix can be defined. The rank condition is relaxed with a $W_X^{collapsed}$, such that $rank(W_X^{collapsed}) = K(T(T+1)/2 - 3) < N$.

It is important to state that the proposed augmented-LSDV estimator does not provide consistent and unbiased estimates of ϕ , but is a projection method that circumvents the nuisance parameters, β and variance-ratio, on the estimate of ϕ .

4.3.2 Recovery of β

In general, the coefficients on the exogenous regressors are of interest and not just lagged dependent parameter. Since the augmented-LSDV method evacuates β from the estimation of ϕ , the point estimates of and inference on β can be recovered by setting the following relationship,

$$z^* = (y^* - y_{-1}^* \phi) = x^* \beta + \epsilon^*.$$

This alternative formulation is no longer dynamic (in the same way as Equation 4.3) so the estimate of β can be obtained in the usual cross-section LSDV framework,

$$\hat{\beta} = (x^{*'} x^*)^{-1} x^{*'} z^*.$$

Since ϕ is unknown, it must be replaced with a consistent and unbiased estimate (discussed below), denoted $\dot{\phi}$, which leads to $z^*(\dot{\phi}) = y^* - y_{-1}^* \dot{\phi}$. The estimate of β conditional on $\dot{\phi}$ can be formulated in a number of ways, specifically

$$\begin{aligned}
\hat{\beta}(\dot{\phi}) &= (x^{*'}x^*)^{-1}x^{*'}z^*(\dot{\phi}) \\
&= \beta + (x^{*'}x^*)^{-1}x^{*'}\epsilon^*(\dot{\phi}) \\
&= (x^{*'}x^*)^{-1}x^{*'}y^* - (x^{*'}x^*)^{-1}x^{*'}y_{-1}^*\dot{\phi} \\
&= (x^{*'}x^*)^{-1}x^{*'}y_{-1}^*\phi + \beta + (x^{*'}x^*)^{-1}x^{*'}\epsilon^* - (x^{*'}x^*)^{-1}x^{*'}y_{-1}^*\dot{\phi} \\
&= \beta + (x^{*'}x^*)^{-1}x^{*'}\epsilon^* + (x^{*'}x^*)^{-1}x^{*'}y_{-1}^*(\phi - \dot{\phi}).
\end{aligned} \tag{4.16}$$

Equation 4.16 is a consistent and unbiased estimate of β , since $E[x^{*'}\epsilon^*] = 0$ and $E[\phi - \dot{\phi}] = 0$. The variance of $\hat{\beta}(\dot{\phi})$ will depend on the variance of $(\phi - \dot{\phi})$, specifically

$$\begin{aligned}
V(\hat{\beta}(\dot{\phi})) &= (x^{*'}x^*)^{-1}x^{*'}V(\epsilon^*)x^*(x^{*'}x^*)^{-1} \\
&\quad + (x^{*'}x^*)^{-1}x^{*'}y_{-1}^*V[\phi - \dot{\phi}]y_{-1}^*x^*(x^{*'}x^*)^{-1} \\
&\quad + 2(x^{*'}x^*)^{-1}E[x^{*'}\epsilon^*]E[\phi - \dot{\phi}]E[y_{-1}^*x^*](x^{*'}x^*)^{-1},
\end{aligned} \tag{4.17}$$

where the last term in this equation will be zero, in expectation. A symmetric confidence set can be constructed from the above based on this variance and assuming asymptotic normal parameter estimates.

An alternative approach is to build the confidence set for ϕ and β . Let $I(\phi, \alpha_\phi)$ represent the confidence interval for ϕ at level α_ϕ . Let $J(\beta(\phi), \alpha_\beta|\phi)$ represent the confidence interval for β at level α_β conditional on ϕ . So the confidence set for ϕ and β is constructed as

$$K(\alpha_\phi, \alpha_\beta) = \{(\phi, \beta) : \dot{\phi} \in I(\phi, \alpha_\phi) \text{ and } \hat{\beta}(\dot{\phi}) \in J(\beta(\phi), \alpha_\beta|\phi)\}. \tag{4.18}$$

This approach is better suited in the context of irregular confidence intervals: asymmetric, disjointed, or empty sets.

4.4 Augmented-LSDV Invariance

Phillips and Sul (2007) derived the analytical bias of the LSDV estimator for panel-ARX(1,K) model and show that the estimate of ϕ depends on the correlation of y_{-1} with the exogenous regressors. An important property of the augmented-LSDV estimator is the partialling out of the exogenous regressors leads to an estimate of ϕ that is free of β . The estimate is distributionally equivalent to a panel-AR(1) model with the same parameter and distribution assumptions on the disturbances.

As discussed in Khalaf and Saunders (2014), σ_γ cannot be identified jointly with the variance of ϵ , so a discontinuous parametric setting for σ_γ is imposed. The inclusion of exogenous regressors extends the initial value assumptions of Khalaf and Saunders (2014), where σ_γ is a function of the variance of the series.

Assumption 6 (Discontinuous σ_γ) *Let the initial value of the series for any i under assumption 4 be a function of $\sigma_\gamma(\phi)$, which is defined as*

$$\sigma_\gamma := \sigma_\gamma(\phi) = \begin{cases} \frac{1}{\sqrt{(1-\phi^2)}} & \text{if } |\phi| < 1, \\ 0 & \text{if } \phi = 1. \end{cases}$$

Under assumption 6, the initial observation is discontinuous in ϕ , specifically

$$y_{i,0} := y_{i,0}(\phi) = \begin{cases} \mu_i + x_{i,0}\beta + \frac{1}{\sqrt{(1-\phi^2)}}\epsilon_{i,0} & \text{if } |\phi| < 1, \\ \mu_i + x_{i,0}\beta & \text{if } \phi = 1. \end{cases}$$

This leads to the invariance theorem for a panel-ARX(1,K) model, estimated by the augmented-LSDV estimator.

Theorem 4.4.1 (Scale and Location Invariance) *Under assumptions 4, 5 and 6, the augmented least-square dummy variable estimate of ϕ from a model defined by*

$$y_{i,t} = (1 - \phi)(\mu_i + x_{i,0}\beta) + \phi y_{i,t-1} + x'_{i,t}\beta + \epsilon_{i,t},$$

is scale invariant in terms of σ_ϵ and Σ_x , and location invariant in terms of μ_i and $x_{i,0}\beta$.

Proof of Theorem 4.4.1 is provided in the Appendix. By Theorem 4.4.1, the augmented-LSDV estimator is invariant to the magnitude of the unobserved heterogeneity, μ_i and $x_{i,0}$, the variance of ϵ and x . As shown in the previous sections the augmented-LSDV estimator is not a function β and hence invariant to β .

The corollaries in Khalaf and Saunders (2014) are valid in the context of the augmented-LSDV estimator and are summarized here.

Corollary 3 *The augmented-LSDV estimator is invariant to affine transformations of the data of the form $\tilde{y}_{i,t} = (\kappa y_{i,t} + \delta_i)$.*

Corollary 3 states that $\tilde{\phi} = \hat{\phi}$, where $\tilde{\phi}$ is the augmented-LSDV estimator from affine transformed data ($\tilde{y}_{i,t} = \kappa y_{i,t} + \delta_i$ and $\kappa \neq 0$), and $\hat{\phi}$ is the augmented-LSDV estimator from the untransformed data ($y_{i,t}$).

Corollary 4 *The LSDV estimator is invariant to σ_γ for $\phi = 1$.*

The implication of corollary 4 is that the augmented-LSDV estimate, $\hat{\phi}$, is not a function of σ_γ , so it is unchanged for any value of σ_γ . Corollaries 3 and 4 are exploited from a test inversion perspective in the proposed method presented below.

4.5 Indirect Inference Estimation: ϕ and β

To obtain consistent parameter estimator for dynamic panel models with exogenous regressors, the literature has focused on the GMM family of estimators. I present here the extension of indirect inference estimation to allow for exogenous regressors. As discussed in Section 4.3.1, several of the bias-correction term methods could incorporate exogenous regressors by implementing the augmented-LSDV estimator for the initial estimate of ϕ . In this section, I extend the indirect inference estimator

(IIE) for dynamic panels in Gouriéroux, Phillips, and Yu (2010) to include exogenous regressors through the use of the augmented-LSDV estimator.

The indirect inference estimator (IIE) is a simulation-based method introduced by Gourieroux, Monfort, and Renault (1993) and independently by Smith (1993) and has been primarily used for bias correction. Gouriéroux, Phillips, and Yu (2010) extend the IIE to the dynamic panel context and show that their approach generates asymptotically consistent estimators of ϕ . Like most simulation methods, IIE has multiple stages that are iterated. The inclusion of exogenous regressors was very briefly discussed in Gouriéroux, Phillips, and Yu (2010, pg. 74), and incorrect as discussed in relation to Equation 4.6.

I present three alternative methods of implementing IIE to a dynamic panel model with exogenous regressors. The first method, denoted IILSDV2, assumes that the IIE objective function depends on two parameters $\theta_2 = (\phi, \beta')'$. The second method, denoted IILSDV3, assumes that the IIE objective function depends on three parameters $\theta_3 = (\phi, \beta', \sigma_\epsilon^2)'$. The third method, denoted aug-IILSDV, uses the augmented-LSDV framework developed above to obtain an estimator of ϕ , which is used to recover estimators of β , so $\theta_{aug} = (\phi)$ and $\beta(\phi)$.

I expect that the IILSDV2 method will erroneously assume scale invariance by naively setting $\sigma_\epsilon^2 = 1$. The IILSDV3 addresses this oversight directly by allowing σ_ϵ^2 to be a free parameter. The augmented-LSDV estimator is scale invariant so σ_ϵ^2 does not affect the estimates of ϕ and the recovery of β .

The indirect inference estimation method is similar for all three methods. As such all will be described by the general framework below, and to illustrate key points the methods are indicated individually. Denote the matrix dependent variables as $w^* = (y_{-1}^*, x^*)$, which leads to the estimators of θ from the data, by method,

$$\hat{\theta}_0 = \begin{cases} (w^{*'}w^*)^{-1} w^{*'}y^* & \text{for IILSDV2} \\ \left((\hat{\phi}_0, \beta'_0)' = (w^{*'}w^*)^{-1} w^{*'}y^*, \hat{\sigma}_{\epsilon,o}^2 = V(\hat{\epsilon}_0^*) \right)' & \text{for IILSDV3} \\ (y_{-1}^{*'}M_X y_{-1}^*)^{-1} y_{-1}^{*'}M_X y^* & \text{for aug-IILSDV} \end{cases} \quad (4.19)$$

Draw as set of error terms, $\epsilon_{i,t,h}$ and $\mu_{i,h}$,

$$\mu_{i,h} \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma_{\mu,h}^2) \quad (4.20)$$

$$\epsilon_{i,t,h} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\epsilon,h}^2) \quad (4.21)$$

where $h = 1, \dots, H$. Independence of the error terms does not affect the result and as such could be relaxed. The normal distribution is used in the current work, but alternative distributions could be assumed if it is suspected that the true model has non-normal disturbances. The assumption about the variance of $\epsilon_{i,t,h}$ is different for each method; both IILSDV2 and aug-IILSDV assume unity, and $\sigma_{\epsilon,h}^2$ is an estimated parameter in the IILSDV3 method.

Construct a set of simulated data based on the null model for each method with a given parameter set, θ_s ,

$$y_{i,t,h} = \begin{cases} (1 - \phi_s) \mu_{i,h} + \phi_s y_{i,t-1,h} + x'_{i,t} \beta_s + \epsilon_{i,t,h} & \text{for IILSDV2} \\ (1 - \phi_s) \mu_{i,h} + \phi_s y_{i,t-1,h} + x'_{i,t} \beta_s + \epsilon_{i,t,h} & \text{for IILSDV3} \\ (1 - \phi_s) \mu_{i,h} + \phi_s y_{i,t-1,h} + \epsilon_{i,t,h} & \text{for aug-IILSDV} \end{cases} \quad (4.22)$$

The simulated series for the aug-IILSDV does not include the exogenous regressors, since the projection matrix (M_X) will partial them out. The IIE methodology allows for differences between the data model and simulation model.

For each of the simulated data series compute the associated parameter estimates, $\hat{\theta}_{s,h}$,

$$\hat{\theta}_{s,h} = \begin{cases} (w_h^{*'} w_h^*)^{-1} w_h^{*'} y_h^* & \text{for IILSDV2} \\ \left((\hat{\phi}_h, \beta_h')' = (w_h^{*'} w_h^*)^{-1} w_h^{*'} y_h^*, \hat{\sigma}_{\epsilon,h}^2 = V(\hat{\epsilon}_h^*) \right)' & \text{for IILSDV3} \\ (y_{-1,h}^{*'} y_{-1,h}^*)^{-1} y_{-1,h}^{*'} y_h^* & \text{for aug-IILSDV} \end{cases} \quad (4.23)$$

where $w_h^* = (y_{-1,h}^*, x^*)$. The simulated parameter estimates are $\hat{\theta}_{s,h} \in \hat{\Theta}_{s,H}$.

The indirect inference estimator of θ is given by

$$\hat{\theta} = \arg \min_{\theta_s \in \Theta} \{Q(\theta_s)\} \quad (4.24)$$

where

$$Q(\theta_s) = \left\| \hat{\theta}_0 - b_T(\hat{\Theta}_{s,H}) \right\| \quad (4.25)$$

and $\|\cdot\|$ represents a measure of the distance. Following the methodology of Gouriéroux, Monfort, and Renault (1993), let $b_T(\hat{\Theta}_{s,H})$ represent a *binding function* that maps the set of simulated parameter estimates to the data-based estimate.

Cconsistency requires that the binding function must be locally injective at the true value of ϕ , refer to Gouriéroux and Monfort (1995). This allows for estimation of θ using simulation, which relaxes the requirement that the binding function is analytically tractable, invertible, or continuous. Defining the binding function as

$$b_T(\hat{\Theta}_{s,H}) = E[\hat{\theta}_{s,h}],$$

allows for consistent estimation as $H \rightarrow \infty$ by the empirical average from the simulated series,

$$b_T(\hat{\Theta}_{s,H}) = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_{s,h}.$$

This leads to the indirect inference objective function,

$$Q(\theta_s) = \left\| \hat{\theta}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_{s,h} \right\|.$$

In keeping with Gouriéroux, Phillips, and Yu (2010), the distance metric is a sandwich form with the identity matrix as the weighting matrix:³

$$Q(\theta_s) = \left(\hat{\theta}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_{s,h} \right)' \left(\hat{\theta}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_{s,h} \right). \quad (4.26)$$

---

³Alternatives include: the Euclidean norm, and the absolute value of the distance. The absolute value may have desirable properties if the distribution of $\hat{\phi}_{s,h}$ is heavily skewed (asymmetric).

In the IIE framework, use of an auxiliary model is allowed, so long as the necessary conditions of the binding function are met. In some cases, an auxiliary model can be simulated with greater ease than the original model applied to the data. Invariance of the augmented-LSDV estimator ensures that the binding function is injective, and the parameter reduction allows for the use of a panel-AR(1) as the auxiliary model.

4.5.1 Bias Correction: Monte Carlo Evidence

A small Monte Carlo simulation study is conducted to assess the bias correction of a small set of approaches. The simulations are based on the following panel-ARX(1,1) model where the exogenous regressor is predetermined, specifically,

$$\begin{aligned} y_{i,t} &= (1 - \phi) (\mu_i + x_{i,0}\beta) + \phi y_{i,t-1} + x_{i,t}\beta + \epsilon_{i,t} \\ x_{i,t} &= (1 - \gamma) \tau_i + \gamma x_{i,t-1} + v_{i,t}. \end{aligned}$$

The parameters of the model are set to $\beta = 0.4$, $\gamma = 0.2$, and over a range of $\phi = \{0.5, 0.8, 0.999\}$, and the errors are drawn from the normal distribution as follows,

$$\begin{aligned} \mu_i &\overset{i.i.d.}{\sim} \mathcal{N}(0, 16) \quad \text{and} \quad \epsilon_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, 4), \text{ and} \\ \tau_i &\overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \quad \text{and} \quad v_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1). \end{aligned}$$

The number of cross-sections and time periods are $N = 200$ and $T = 6$, respectively. The initial values for each series are set to $y_{i,0} = \mu_i + x_{i,0}\beta$ and $x_{i,0} = \tau_i$, which exploits the location invariance conditional on $\sigma_\gamma = 0$ (Theorem 4.4.1).

Table 4.1 presents the bias of each method. As expected the LSDV approach is heavily biased, and primarily used as a frame of reference. The system GMM method of Blundell and Bond (1998) provides good bias correction, but this deteriorates for processes with less persistence. This is a direct consequence of a high variance ratio ($\sigma_\mu^2/\sigma_\epsilon^2 = 4$), which conforms with the findings of Bun and Windmeijer (2010).

Table 4.1: Parameter Estimate Bias for a Panel-ARX(1,1) Model

Lagged Dependent	$\phi = 0.5$	$\phi = 0.8$	$\phi = 0.999$
LSDV	-0.3541	-0.4963	-0.5599
System GMM	0.1181	0.0264	-0.0013
GPY	-0.1099	-0.2506	-0.3231
IILSDV2	-0.0600	-0.0909	-0.0947
IILSDV3	-0.0072	-0.0117	-0.0078
aug-IILSDV	-0.0048	-0.0004	-0.0119
Exogenous	$\beta = 0.4$		
LSDV	-0.0233	-0.0604	-0.0950
System GMM	0.0033	-0.0183	0.0015
GPY	-0.0045	-0.0366	-0.0825
IILSDV2	0.0068	-0.0184	-0.0500
IILSDV3	0.0082	0.0036	0.0025
aug-IILSDV	-0.0002	0.0000	-0.0028

Based on 1000 Monte Carlo simulations. All calculations performed in GAUSS except system GMM which used STATA 12.0 with the two-step and robust errors (Windmeijer (2005)).

The indirect inference LSDV method (IILSDV2) that conditions on the variance of the time varying error term to be unity ($\sigma_\epsilon^2 = 1$) is downward biased, since the LSDV estimator is not scale invariant when exogenous regressors are present. The indirect inference LSDV method (IILSDV3) that uses all free parameters in the objective function leads to good bias correction performance. The indirect inference method that employs the augmented-LSDV approach (aug-IILSDV) also has good bias correction performance.

The methods presented in Table 4.1 provide suitable bias correction when compared to the LSDV parameter estimate for the exogenous regressors. Notably, the augmented-LSDV method results in near unbiased estimates, while the bias correction of the GPY and IILSDV2 methods deteriorate as the persistence of the series increases.

4.6 Confidence Set for ϕ

The approach of Khalaf and Saunders (2014) is extended to incorporate exogenous regressors and introduce the augmented-LSDV to compute exact confidence sets for the parameters. As demonstrated in the previous section, the inclusion of exogenous regressors is not a simple extension of the indirect inference method for dynamic panels to obtain consistent parameter estimates.

Bias correction has commanded most of the focus in dynamic panels models, with limited research into the efficiency of the estimators. The estimators described above depend on asymptotics to obtain the confidence set of the parameter estimates and, in most of the cases, the limiting distribution is normal.

This paper extends the indirect confidence set inference (ICSI) method for dynamic panels, developed in Khalaf and Saunders (2014), to allow for the inclusion of exogenous regressors. The use of the augmented-LSDV ensures that the IIE generates exact confidence set estimates by incorporating the Monte Carlo method of Dufour (2006).

The proposed procedure inverts the IIE objective function, Equation 4.26, to build a confidence set for ϕ . To begin, I estimate ϕ from the data using the augmented-LSDV estimator and denote it as $\hat{\phi}_0$. Next, I impose the joint null,

$$H_s : \phi = \phi_s \text{ and } \sigma_\gamma = \sigma_\gamma(\phi_s) \quad (4.27)$$

over a reasonable range for ϕ_s , specifically $\phi_s \in (-1, 1]$, and assumption 6 defines the parametric function for $\sigma_\gamma(\phi_s)$.

A set of H simulation draws of the error terms are used to generate a set of simulated data conditioned on ϕ_s given by

$$y_{i,0,h|s} = \sigma_\gamma \epsilon_{i,0,h} \quad (4.28)$$

$$y_{i,t,h|s} = \phi_s y_{i,t-1,h|s} + \epsilon_{i,t,h}. \quad (4.29)$$

Location invariance allows for the selection of $\mu_{i,h} = 0 \forall i$, which simplifies the simulations. The use of a simulation model that excludes regressors is appropriate by theorem 4.4.1, since the estimators share the same limiting distribution. The LSDV estimate from each of these can be obtained in a similar manner as above

$$\hat{\phi}_{h|s} = (y_{-1,h|s}^{*'} y_{-1,h|s}^*)^{-1} y_{-1,h|s}^{*'} y_{h|s}^* \quad (4.30)$$

The binding function is estimated using the sample mean from the simulated series,

$$b_T(\hat{\Phi}_{H|s}) = \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s}. \quad (4.31)$$

The objective function derived from the data is

$$Q(\phi_{0|s}) = \left(\hat{\phi}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right)' \left(\hat{\phi}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right). \quad (4.32)$$

Theorem 4.6.1 (Objective Function Location-Scale Invariant) *Under assumptions 4, 5 and 6, and the joint null, defined as $H_s : \phi = \phi_s$ and $\sigma_\gamma = \sigma_\gamma(\phi_s)$, the distribution of $Q(\phi_s)$ is completely determined by ϕ_s and all the disturbances.*

This means that the distribution of $Q(\phi_s)$ is pivotal, or nuisance parameter free, under the null.

Next, draw M Monte Carlo replications and generate the dependent variable conditional on ϕ_s , similar to the procedure used for the H simulated series. The resulting replications and LSDV estimates are

$$\begin{aligned} y_{i,t,m} &= \phi_s y_{i,t-1,m} + \epsilon_{i,t,m} \\ \hat{\phi}_{m|s} &= (y_{-1,m}^{*'} y_{-1,m}^*)^{-1} y_{-1,m}^{*'} y_m^*. \end{aligned}$$

Again, location invariance allows for the assumption that $\mu_{i,m} = 0 \forall i$.

A set of Monte Carlo objective functions are computed, denoted as $Q_M(\phi_s)$,

$$Q(\phi_{m|s}) = \left(\hat{\phi}_{m|s} - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right)' \left(\hat{\phi}_{m|s} - \frac{1}{H} \sum_{h=1}^H \hat{\phi}_{h|s} \right). \quad (4.33)$$

Any individual Monte Carlo replication ($\hat{\phi}_{m|s}$) and a given simulated series ($\hat{\phi}_{h|s}$) are exchangeable, since both were constructed under the same joint null.

The p -value is obtained by extending the method given in Dufour (2006). First, count the number elements in $Q_M(\phi_s)$ that exceed $Q(\phi_{0|s})$,

$$\text{count}(Q(\phi_{0|s})) = \sum_{m=1}^M \mathbf{1}(Q(\phi_{0|s}) \leq Q(\phi_{m|s})), \quad (4.34)$$

and $\mathbf{1}(\cdot)$ is one when the statement is true and zero otherwise. The p -value is computed as

$$\text{p-value}(Q(\phi_{0|s})) = \frac{\text{count}(Q(\phi_{0|s})) + 1}{M + 1}. \quad (4.35)$$

The confidence set for ϕ at significance level α is constructed by collecting all the values of ϕ that satisfy the right-tailed condition

$$CI(\phi|\alpha) = \{\phi_s \mid \text{p-value}(Q(\phi_{0|s})) > \alpha\} \quad (4.36)$$

The indirect confidence set inference method has a number of desirable properties. The confidence set will be exact in the sense that

$$Pr(\phi \in CI(\phi|\alpha)) = 1 - \alpha, \quad (4.37)$$

since the distribution of ϕ is free of nuisance parameters. Since the H simulated and M Monte Carlo series are exchangeable under the null, the confidence set will have level control. It is required that the distribution of the IIE objective function can be simulated, but the distribution can be unknown.

4.7 Simulation Study

I conduct a simulation study to examine the coverage properties of the indirect confidence set inference method for a panel-ARX(1,1) model. The data generating process for this study is defined by the following set of equations:

$$\begin{aligned} y_{i,0} &= x_{i,0}\beta + \sigma_\gamma \epsilon_{i,0} \\ y_{i,t} &= (1 - \phi) x_{i,0}\beta + \phi y_{i,t-1} + x_{i,t}\beta + \epsilon_{i,t} \\ x_{i,0} &= v_{i,0} \left(\frac{1}{\sqrt{1 - \gamma^2}} \right) \\ x_{i,t} &= \gamma x_{i,t-1} + v_{i,t}. \end{aligned}$$

I exploit location invariance of the augmented-LSDV estimator to avoid unnecessary parameter settings, specifically $\mu_i = 0$ for all i . The coefficient on the exogenous regressor is $\beta = 0.6$, and the lagged dependent parameter is examined over various degrees of persistence, $\phi = \{0.0, 0.6, 0.9, 0.999, 1\}$. The dependent series is generated from draws of ϵ_i from the standard normal distribution. The exogenous regressor is stationary with a zero cross-section mean, so $\gamma = 0.4$ and v_i is drawn from the standard normal distribution.

The rejection frequency of the indirect confidence set inference method is tested at a range of points, $\phi_s = \{[-0.95, 0.95], 0.999, 1.0\}$, and the augmented-LSDV does not require setting β under the null. The number of cross-sections is modest, $N = \{100, 200, 300\}$, and the number of time periods is kept low, $T = \{5, 10, 20\}$. The number of simulated series for the indirect inference objective functions is $H = 20$ and the number of Monte Carlo indirect inference objective functions is $M = 79$. The rejection frequency is based on 1000 replications.

The main simulation results are presented in Tables 4.2 – 4.6. The rejection frequency under the null is close to the 5% level and ranges from 3.8% to 5.8%. As expected, the simulations demonstrate that the method will have level control, in probability. The rejection frequencies under the alternative increase as ϕ deviates

from the true parameter value, and improve as either the number of cross-sections (N) or the number of time periods (T) increase. At the unit root, $\phi = 1$, the method is level correct under the null and demonstrates good power properties as ϕ deviates from unity. Comparing the power of $\phi = 0.999$ and $\phi = 1$, Tables 4.5 and 4.6 show that the discontinuous null has little effect on the rejection frequencies, so unity in the proposed framework is *just another point* to be considered.

The proposed method is subjected to several robustness tests on the data generating process. Results are presented in Table 4.7 where column A is taken from Table 4.3 and used as the benchmark. Increasing the number of simulated series (column C and $H = 40$) and Monte Carlo objective function realizations (column B and $M = 199$) has almost no effect on the rejection frequencies under the null, but some small changes under the alternative. Changing the data generating process for the exogenous regressor (columns D and E and $\gamma = \{0, -0.4\}$) results in marginal changes in the rejection frequencies. When the exogenous regressor is not in the model, $\beta = 0$, the rejection frequencies in column F are identical to the benchmark column A . This also demonstrates that the augmented-LSDV estimator removes not only the exogenous regressor but also any inadvertent correlation. The collapsed- M_X matrix is used to relieve some degrees of freedom, the rejection frequencies in column G are similar to column A . Column H imposes a model without exogenous regressors and does not require the augmented-LSDV estimator, and the rejection frequencies are similar to column A . This indicates that the augmented-LSDV estimator is equivalent to using a collapsed version of the projection matrix and a simple panel-AR(1) model. However, the collapsing and the reduction to a simple panel-AR(1) model relaxes to some extent the $N > TK$ restrictions.

4.8 Concluding Remarks

The augmented-LSDV estimator derived in this paper is unique to dynamic panel since the number of cross-sections is large relative to the number of time periods. The

proposed approach leads to nuisance parameter free estimators of parameters and associated statistics, with respect to exogenous regressors. The resulting estimator is equivalent to the panel-AR(1) model, and in the literature this simpler model has been well analyzed leading to several consistent bias-correction methods. The augmented indirect inference estimator adds to the available bias-correction methods, and the objective function allows for inversion to obtain exact confidence sets for the parameters.

The augmented-LSDV estimator has a rank condition, $N > TK$, that could be construed as restrictive for data sets with a fixed number of cross-sections, like country or state panels. Easing the rank constraint on the augmented-LSDV is viewed as a worthy extension of the method, and would expand the scope of suitable data sets.

The augmentation matrix (M_X) is a general approach that can be applied to existing approaches, specifically bias-correction term methods, GMM and IV methods, and X-differencing. In some of these methods the objective functions are location-scale invariant, so the application of the ICSI method would result in exact confidence sets.

[illegible]

Table 4.3: Rejection Frequency: $\phi = 0.6$ at 5% Significance Level

	N=100			N=200			N=300		
ϕ	5	10	20	5	10	20	5	10	20
0.1	1	1	1	1	1	1	1	1	1
0.15	0.998	1	1	1	1	1	1	1	1
0.2	0.995	1	1	1	1	1	1	1	1
0.25	0.978	1	1	1	1	1	1	1	1
0.3	0.919	1	1	0.998	1	1	1	1	1
0.35	0.792	1	1	0.976	1	1	0.997	1	1
0.4	0.584	0.996	1	0.898	1	1	0.97	1	1
0.45	0.376	0.915	1	0.664	1	1	0.821	1	1
0.5	0.19	0.647	0.957	0.326	0.894	1	0.488	0.981	1
0.55	0.083	0.186	0.468	0.125	0.357	0.82	0.145	0.521	0.943
0.6	0.058	0.043	0.051	0.041	0.043	0.038	0.042	0.044	0.033
0.65	0.084	0.186	0.478	0.105	0.356	0.816	0.136	0.492	0.947
0.7	0.185	0.602	0.971	0.289	0.903	1	0.44	0.979	1
0.75	0.347	0.926	1	0.61	0.996	1	0.788	1	1
0.8	0.521	0.992	1	0.838	1	1	0.957	1	1
0.85	0.702	0.999	1	0.962	1	1	0.999	1	1
0.9	0.86	1	1	0.994	1	1	1	1	1
0.95	0.938	1	1	1	1	1	1	1	1
0.999	0.978	1	1	1	1	1	1	1	1
1	0.977	1	1	1	1	1	1	1	1

Table 4.4: Rejection Frequency: $\phi = 0.9$ at 5% Significance Level

	N=100			N=200			N=300		
ϕ	5	10	20	5	10	20	5	10	20
0.4	1	1	1	1	1	1	1	1	1
0.45	0.997	1	1	1	1	1	1	1	1
0.5	0.983	1	1	1	1	1	1	1	1
0.55	0.938	1	1	0.997	1	1	1	1	1
0.6	0.86	1	1	0.993	1	1	0.999	1	1
0.65	0.705	0.999	1	0.956	1	1	0.994	1	1
0.7	0.5	0.98	1	0.851	1	1	0.937	1	1
0.75	0.301	0.884	1	0.579	0.998	1	0.768	1	1
0.8	0.164	0.56	0.97	0.285	0.837	1	0.409	0.954	1
0.85	0.079	0.168	0.513	0.101	0.325	0.866	0.126	0.47	0.967
0.9	0.054	0.046	0.053	0.036	0.049	0.038	0.044	0.054	0.04
0.95	0.078	0.183	0.526	0.098	0.349	0.863	0.133	0.449	0.967
0.999	0.163	0.539	0.981	0.272	0.84	1	0.392	0.963	1
1	0.163	0.553	0.984	0.274	0.845	1	0.397	0.962	1

Table 4.5: Rejection Frequency: $\phi = 0.999$ at 5% Significance Level

	N=100			N=200			N=300		
ϕ	5	10	20	5	10	20	5	10	20
0.4	1	1	1	1	1	1	1	1	1
0.45	1	1	1	1	1	1	1	1	1
0.5	0.999	1	1	1	1	1	1	1	1
0.55	0.996	1	1	1	1	1	1	1	1
0.6	0.976	1	1	0.999	1	1	1	1	1
0.65	0.939	1	1	0.998	1	1	1	1	1
0.7	0.846	1	1	0.988	1	1	0.999	1	1
0.75	0.671	1	1	0.948	1	1	0.989	1	1
0.8	0.474	0.98	1	0.82	1	1	0.934	1	1
0.85	0.291	0.857	0.999	0.541	0.991	1	0.747	1	1
0.9	0.151	0.505	0.971	0.254	0.807	1	0.399	0.946	1
0.95	0.075	0.153	0.514	0.108	0.288	0.848	0.132	0.441	0.947
0.999	0.04	0.043	0.044	0.041	0.057	0.044	0.047	0.053	0.056
1	0.041	0.045	0.046	0.042	0.054	0.046	0.048	0.061	0.054

Table 4.6: Rejection Frequency: $\phi = 1$ at 5% Significance Level

	N=100			N=200			N=300		
ϕ	5	10	20	5	10	20	5	10	20
0.45	1	1	1	1	1	1	1	1	1
0.5	0.999	1	1	1	1	1	1	1	1
0.55	0.996	1	1	1	1	1	1	1	1
0.6	0.977	1	1	0.999	1	1	1	1	1
0.65	0.936	1	1	0.998	1	1	1	1	1
0.7	0.844	1	1	0.989	1	1	0.999	1	1
0.75	0.678	1	1	0.948	1	1	0.989	1	1
0.8	0.48	0.983	1	0.818	1	1	0.94	1	1
0.85	0.301	0.857	1	0.546	0.993	1	0.752	1	1
0.9	0.148	0.505	0.974	0.255	0.807	1	0.402	0.948	1
0.95	0.074	0.159	0.526	0.107	0.294	0.855	0.137	0.452	0.955
0.999	0.04	0.046	0.046	0.044	0.054	0.05	0.047	0.057	0.052
1	0.041	0.045	0.045	0.043	0.055	0.052	0.047	0.058	0.051

Table 4.7: Rejection Frequency: $\phi = 0.6$, $N = 200$, and $T = 5$ at 5% Significance Level

	A	B	C	D	E	F	G	H
	M_X	M_X	M_X	M_X	M_X	M_X	collapsed-M_X	$I_{N(T-1)}$
β	0.6	0.6	0.6	0.6	0.6	0	0.6	0
γ	0.4	0.4	0.4	0	-0.4	0.4	0.4	—
H	20	20	40	20	20	20	20	20
M	79	199	79	79	79	79	79	79
0.2	1	1	1	1	1	1	1	1
0.25	1	0.999	1	1	1	1	1	1
0.3	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999
0.35	0.976	0.98	0.979	0.974	0.976	0.976	0.978	0.978
0.4	0.898	0.898	0.896	0.9	0.902	0.898	0.896	0.904
0.45	0.664	0.677	0.663	0.662	0.664	0.664	0.662	0.663
0.5	0.326	0.327	0.326	0.327	0.333	0.326	0.325	0.329
0.55	0.125	0.112	0.122	0.128	0.124	0.125	0.128	0.125
0.6	0.041	0.044	0.049	0.043	0.042	0.041	0.042	0.042
0.65	0.105	0.102	0.107	0.102	0.102	0.105	0.106	0.108
0.7	0.289	0.301	0.303	0.296	0.297	0.289	0.301	0.316
0.75	0.61	0.623	0.61	0.608	0.612	0.61	0.611	0.619
0.8	0.838	0.846	0.851	0.835	0.839	0.838	0.837	0.857
0.85	0.962	0.964	0.964	0.963	0.964	0.962	0.961	0.967
0.9	0.994	0.996	0.994	0.994	0.992	0.994	0.994	0.994
0.95	1	1	1	1	1	1	1	1

5. DYNAMIC TECHNICAL EFFICIENCY

Charles J. Saunders and Lynda Khalaf

Similar version: Proceedings of the North American Productivity Workshop 2014

5.1 Introduction

Several methods have been proposed to correct for the bias of dynamic panel data models including, but not limited to, GMM, bias-correction terms, indirect inference, and X-differencing. For some approaches, there are relatively few restrictions on the distributions of the errors, commonly Gaussian errors are assumed. Some dynamic panel approaches allow for less restrictive distributions, for example the skew-normal distribution. Relatively few studies take into account the possibility of dynamics in the panel structure and non-Gaussian errors; see Ahn and Sickles (2000), Desli et al. (2003) and Tsionas (2006).

The first contribution of this paper demonstrates that the inclusion of dynamics into a stochastic frontier model can lead to biased parameter estimators. Maximum likelihood estimators that are designed to take into account the incidental parameter bias are few. The implications of this bias for a MLE approach is examined via simulations, which demonstrates that the parameter bias is large. The MLE parameter bias has little effect on the estimation of the distribution parameter related to skewness.

The second contribution is the introduction of the indirect inference estimation (IIE) approach for bias correction. The approach extends that of Gouriéroux et al. (2010) to incorporate the additional parameters by allowing for skew-Normal errors. This extension leads to bias correction of the lagged dependent parameter and relatively unbiased skewness parameter estimators. The confidence intervals for the pa-

rameters are constructed with the indirect confidence set inference (ICSI) approach of Khalaf and Saunders (2014) enhanced to consider non-Gaussian distributions. Simulations show that the IIE leads to relatively unbiased parameter estimates, and the ICSI method leads to good coverage properties both under the null and alternative.

Commercial bank costs are analyzed in a dynamic technical efficiency model framework. A negative dynamic relationship is indicated by MLE, while IIE provides evidence that a static model is more appropriate. The confidence intervals are constructed under the ICSI method, which show that the model errors have a positive skew.

The remainder of the chapter is structured as follows. Section 2 presents the current models available in the literature and presents the dynamic stochastic frontier model introduced in this chapter. Section 3 presents the MLE approach and introduces the IIE and ICSI methods. Section 4 covers the Monte Carlo simulation study, which demonstrates the MLE-bias and IIE bias-correction and the coverage properties of the ICSI method to construct confidence sets. Section 5 examines cost inefficiency of commercial banks under the assumption of dynamic technical inefficiency. Section 6 summarizes and provides direction to future research.

5.2 Methods

The dynamic technical efficiency model put forth in Ahn and Sickles (2000) is defined by

$$y_{i,t} = x_{i,t}\beta + (\beta_0 + \gamma t) + v_{i,t} - u_{i,t} \quad (5.1)$$

$$u_{i,t} = (1 - \rho_i)u_{i,t-1} + (1 - \rho_i)\gamma + \eta_{i,t}, \quad (5.2)$$

where $v_{i,t}$ are i.i.d. symmetric errors, $\eta_{i,t} \geq 0$, and $0 < \rho_i \leq 1$, such that $u_{i,t} \geq 0$. To take into account the serial correlation of the errors, the model can be re-specified as:

$$y_{i,t} = (1 - \rho_i)y_{i,t-1} + (x_{i,t} - (1 - \rho_i)x_{i,t-1})\beta \quad (5.3)$$

$$+ \rho_i\beta_0 + \rho_i\gamma t + (v_{i,t} - (1 - \rho_i)v_{i,t-1}) - \rho_i\eta_{i,t}. \quad (5.4)$$

Clearly the lagged dependent series is correlated with $v_{i,t-1}$, so the authors suggest two solutions for estimation. The first assumes that $v_{i,t} = (1 - \rho_i)v_{i,t-1} + h_{i,t}$ where $h_{i,t}$ are i.i.d. symmetric errors centered at zero, and the parameters can be estimated via non-linear GLS. The second uses appropriate lags of the dependent series as instruments to estimate the parameters by non-linear GMM.

The authors test the assumption that individual efficiency dynamics are equal, $\rho_i = \rho$ for all i , which places the model closer to the standard dynamic panel frameworks of Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1998). The GMM approach provides consistent parameter estimators, but more recent studies have found that these estimators are susceptible to weak instruments, especially near the unit root boundary (see Phillips (2014b)), and a variance ratio of the errors and unobserved heterogeneity that is not unity (see Bun and Windmeijer (2010)).

A second representation of the dynamic technical efficiency model put forth in Desli et al. (2003) is

$$y_{i,t} = x_{i,t}\beta + \alpha_{i,t} + v_{i,t} \text{ and} \quad (5.5)$$

$$\alpha_{i,t} = \alpha_i + \rho\alpha_{i,t-1} + w_{i,t}\gamma - u_{i,t}, \quad (5.6)$$

where $u_{i,t} \geq 0$, and $w_{i,t}$ are a set of technical efficiency specific covariates. The model can be represented as a panel-ARMA, specifically

$$y_{i,t} = \alpha_i + \rho y_{i,t-1} + (x_{i,t} - \rho x_{i,t-1})\beta + w_{i,t}\gamma + \epsilon_{i,t} \text{ and} \quad (5.7)$$

$$\epsilon_{i,t} = (v_{i,t} - \rho v_{i,t-1}) - u_{i,t}. \quad (5.8)$$

This representation will suffer from an incidental parameter bias due to the correlation of the unobserved heterogeneity (α_i) and the lagged dependent parameter ($y_{i,t-1}$).

The third representation of dynamic technical efficiency is that of Tsionas (2006), given as

$$y_{i,t} = x_{i,t}\beta + v_{i,t} - u_{i,t} \text{ and} \quad (5.9)$$

$$\ln(u_{i,t}) = \rho \ln(u_{i,t-1}) + w_{i,t}\gamma + \eta_{i,t}, \quad (5.10)$$

where $v_{i,t} \sim IN(0, \sigma_v^2)$ and $\eta_{i,t} \sim IN(0, \sigma_\eta^2)$. Although less clear by this representation, if any regressor in $x_{i,t}$ is a time-invariant regressor (for example x_i) will be correlated with the lagged errors ($u_{i,t-1}$), since this is effectively introducing heterogeneity.

The dynamic technical efficiency model of both Desli et al. (2003) and Tsionas (2006) will lead to biased parameter estimates. Without loss of generality, this paper focuses on an autoregressive panel without regressors, specifically

$$y_{i,t} = \rho y_{i,t-1} + (1 - \rho)\alpha_i + \eta_{i,t} \text{ and} \quad (5.11)$$

$$\eta_{i,t} \sim \text{skew-}N(\mu, \sigma_\eta, \lambda), \quad (5.12)$$

where μ is the location, σ_η is the scale, and λ is the shape parameter (the normal distribution is recovered when $\lambda = 0$). By transformation, the errors can be alternatively defined by $\eta_{i,t} = v_{i,t} - u_{i,t}$, where $v_{i,t} \sim IN(0, \sigma_v^2)$ and $u_{i,t} \sim IN^+(0, \sigma_u^2)$.

This dynamic technical efficiency model has four parameters of interest, defined as $\theta = \{\rho, \mu, \sigma_\eta, \tau\}$. Dynamic panel models with normal errors lead to provably

location-scale invariant estimates of ρ ; see Khalaf and Saunders (2014). In contrast, panel models with skew-normal errors lead to location-scale invariant estimates of ρ , but these estimates are not shape parameter invariant.

5.3 Unbiased Estimators and Confidence Set

The indirect inference estimator (IIE) produces unbiased estimates of the parameter (θ). The parameter confidence set is constructed by inverting the indirect inference objective function (IIOF) by introducing an additional layer of Monte Carlo replications.

The estimates from the data are obtained by maximum likelihood from the demeaned series,¹

$$\eta_{i,t}^* = y_{i,t}^* - \rho y_{i,t-1}^*, \quad (5.13)$$

where the individual likelihood is given by

$$L(\eta_{i,t}^*; \mu, \sigma_\eta, \lambda) = \frac{2}{\sigma_\eta} \phi\left(\frac{\eta_{i,t}^* - \mu}{\sigma_\eta}\right) \Phi\left(\frac{\lambda(\eta_{i,t}^* - \mu)}{\sigma_\eta}\right), \quad (5.14)$$

and the total log-likelihood is

$$\mathcal{L}(\rho, \mu, \sigma_\eta, \lambda) = \sum_{i=1}^N \sum_{t=2}^T \ln(L((y_{i,t}^* - \rho y_{i,t-1}^*); \mu, \sigma_\eta, \lambda)). \quad (5.15)$$

The MLE parameter estimate of ρ can be shown to be equivalent to the fixed-effect estimator for dynamic panel models. For fixed-T, Nickell (1981) derived the analytical bias for dynamic panel models under the assumption of i.i.d. errors and shows that the bias is $O(T^{-1})$. The i.i.d. errors cover the skew-normal case, but the variance of these errors depends on both the scale (σ_η) and shape (λ) parameters. The structure of the fixed-effect estimator requires that only *strictly* exogenous regressors can be

---

¹The demeaning subtracts the sample mean from the series, and is indicated by an asterisk.

included, but these regressors can be correlated with the unobserved heterogeneity. As shown in Phillips and Sul (2007), the source of the Nickell bias is due to the correlation of the demeaned lagged dependent series with the demeaned errors, and the parameter estimates for the exogenous regressors are subsequently biased due to the original Nickell bias.

The IIE is introduced as a correction approach. Let $\theta = (\rho, \mu, \sigma_\eta, \lambda)'$ represent the vector of parameters. The IIE method constructs a set of simulated series under a given null, $H_0 : \theta = \theta_s$, and each series is indexed by $h = 1, \dots, H$. The IIOF is defined as

$$Q(\theta_{s,0}) = \left(\hat{\theta}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h(\theta_s) \right)' \left(\hat{\theta}_0 - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h(\theta_s) \right), \quad (5.16)$$

where $\hat{\theta}_0$ is the estimate attained from the data. The minimization of $Q(\theta_{s,0})$ over a suitable parameter space, say $\theta_s \in \Theta$, leads to an unbiased and consistent estimator of θ .

The confidence set for θ is constructed using the indirect confidence set inference (ICSI) approach, which generates an additional layer of Monte Carlo simulations that are used to calibrate the IIOF statistic. These Monte Carlo simulations are generated under the joint null, $H_0 : \theta = \theta_s$. The ICSI method uses an M -dimension set of IIOF, specifically

$$Q(\theta_{s,m}) = \left(\hat{\theta}_m - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h(\theta_s) \right)' \left(\hat{\theta}_m - \frac{1}{H} \sum_{h=1}^H \hat{\theta}_h(\theta_s) \right), \quad (5.17)$$

to simulate the empirical distribution of IIOF as a statistic. The rank of $Q(\theta_{s,0})$ in the set of Monte Carlo simulated IIOF is used to compute the Monte Carlo p -value at θ_s . The confidence set is then constructed by collecting all values of θ_s that are not rejected under the null,

$$CI(\theta) = \{\theta_s | \text{p-value}(Q(\theta_{s,0})) > \alpha\}. \quad (5.18)$$

5.4 Simulations

A small simulation study is used to examine the bias correction and the coverage of the confidence set. The data generating process for the simulations is described by Equations 5.11 and 5.12. The simulation study considers static ($\rho = 0$) and stationary ($\rho = 0.4$) panel settings, with symmetric errors ($\lambda = 0$) and positive skew errors ($\lambda = 2$), and for all simulations the scale parameter is fixed, $\sigma_\eta = 1.2$.

Table 5.1: Estimation Bias for a Dynamic Stochastic Frontier Model

			MLE			IIE		
ρ	σ_η	λ	ρ	σ_η	λ	ρ	σ_η	λ
0	1.2	0	-0.249	-0.105	0.011	0.013	0.175	-0.032
0	1.2	2	-0.239	-0.294	-0.721	0.014	0.000	-0.018
0.4	1.2	0	-0.376	-0.139	0.011	0.014	0.196	-0.016
0.4	1.2	2	-0.365	-0.340	-0.810	0.016	0.005	0.039

Based on 1000 replications.

Table 5.1 presents the parameter bias ($\hat{\theta} - \theta$) of each estimation method. The lagged dependent parameter is downward biased for the MLE estimate in all settings and the IIE approach provides suitable bias correction.

The bias of distribution related parameters depends on whether the distribution is Gaussian or not under the null. For normally distributed errors, the MLE and IIE approaches lead to relatively unbiased shape parameter estimates. However, the scale parameter estimates are downward biased for MLE and upward biased for IIE. The IIE estimates under normal errors are provably location-scale invariant (see Khalaf and Saunders (2014)). This is an example of a Davies (1977, 1987) problem; specifically, when the true errors are normally distributed then under the null of normality the scale parameter is not identified.

The IIE parameter estimates are unbiased when the true errors are non-Gaussian, and the scale parameter is identified under both the null and alternative. The MLE parameter estimates are downward biased for all parameters, for positive skew errors.

Tables 5.2 and 5.3 present the rejection frequency properties of the ICSI method for a dynamic technical efficiency model. The former table tests under the null of various values of ρ and the latter for various shape parameter values (the null used in the data generating processes is indicated in bold). The rejection frequency under the null is size corrected with rejection frequencies close to the desired 5% level. The inversion of the HIOF statistic in the ICSI framework ensures that the Monte Carlo and II simulations are exchangeable under the null.

Table 5.2: ICSI Rejection Frequency for Lagged Dependent Parameter

ρ	$\{ \rho = 0.4, \lambda = 0 \}$	$\{ \rho = 0.4, \lambda = 2 \}$
-0.8	0.705	1.000
-0.6	0.386	1.000
-0.4	0.209	0.998
-0.2	0.114	0.939
0.0	0.067	0.418
0.2	0.055	0.104
0.4	0.044	0.050
0.6	0.050	0.081
0.8	0.062	0.221

Based on 1000 replications, $\sigma_\eta = 1.2$, significance level of 5%, and size indicated in bold.

Table 5.3: ICSI Rejection Frequency for Shape Parameter

λ	$\{ \rho = 0.4, \lambda = \mathbf{0} \}$	$\{ \rho = 0.4, \lambda = \mathbf{2} \}$
-5	1.000	1.000
-2	0.978	1.000
-1	0.278	0.998
0	0.044	0.806
1	0.259	0.082
2	0.975	0.050
5	1.000	0.731

Based on 1000 replications, $\sigma_\eta = 1.2$, significance level of 5%, and size indicated in bold.

The rejection frequency under the alternative increases as the parameter setting under the null moves away from the true parameter value. A nonzero shape parameter improves power in ρ and the ICSI approach is robust to asymmetric confidence intervals for λ . The standard MLE-based confidence bounds impose symmetry, which may be a strong assumption in the context of asymmetric distributions.

5.5 Banks, Productivity, and Dynamics

The bank cost data analyzed in Kumbhakar and Tsionas (2005) is used in this study to illustrate the effect of introducing dynamics into a panel model with skew-Normal errors. The data was originally obtained from the Federal Reserve Bank of Chicago, specifically the commercial bank and bank holding company database. The database is balanced with 500 commercial banks identified, and for the years 1996 through to 2000.²

The model under consideration is a dynamic translog model with skew-normal errors, given by:

$$c_{i,t} = c_{i,t-1}\rho + X'_{i,t}\beta + (1 - \rho)\alpha_i + \eta_{i,t} \text{ and} \quad (5.19)$$

$$\eta_{i,t} \sim \text{skew-}N(\mu, \sigma_\eta, \lambda), \quad (5.20)$$

where $c_{i,t}$ is the log costs of the bank, $X_{i,t}$ is the vector of exogenous regressors (including interaction terms), and α_i is the unobserved bank-specific heterogeneity. The exogenous regressors include the quantity and prices of labour, capital, purchased funds, interest-bearing deposits in total transactions accounts, and interest-bearing deposits in total nontransaction accounts.

The exogenous regressors are partialled out to focus on the bias-correction of the lagged dependent parameter, and the parameters of the skew-normal distribution. A

---

²Data set downloaded from the website of William Greene,
<http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm>.

projection matrix, M_X is constructed and under the assumption of *strict* exogeneity.³ The series are demeaned (identified with an asterisk), so the matrix representation of the model is:

$$M_X c^* = M_X c_{-1}^* \rho + M_X \eta^* \text{ and} \quad (5.21)$$

$$M_X \eta^* \sim \text{skew-}N(\mu, \sigma_\eta, \lambda), \quad (5.22)$$

The demeaning of the series forces the sample and expected mean of the estimation errors to be exactly zero. The estimate of μ is obtained recursively by:

$$\hat{\mu} = -\hat{\sigma}_\eta \frac{\hat{\lambda}}{\sqrt{1 + \hat{\lambda}}} \sqrt{\frac{2}{\pi}}, \quad (5.23)$$

since it is exactly defined from the other parameters then it is excluded from the objective functions of the estimators.

Let $\theta = \{\rho, \sigma_\eta, \lambda\}$ represent the model parameters that are estimated by MLE and IIE. The MLE estimators are obtained by maximizing the total likelihood given by:

$$\mathcal{L}(\rho, \sigma_\eta, \lambda) = \sum_{i=1}^N \sum_{t=2}^T \ln \left(L \left((M_X c^* - M_X c_{-1}^* \rho)_{i,t}; \mu(\sigma_\eta, \lambda), \sigma_\eta, \lambda \right) \right). \quad (5.24)$$

The MLE estimates are used as the estimates from the data (θ_0) for the IIE and ICSI methods outlined above.

Table 5.4: Dynamic Translog Model of Commercial Banks

			IIE 95% Confidence Interval	
	MLE	IIE	Lower	Upper
ρ	-0.2340	0.0123	-0.4620	0.4786
σ_η	0.2564	0.3676	<0.00001	0.9052
λ	1.5449	3.1886	0.7536	>15.0000

³On the definition of M_X : Gouriéroux et al. (2010) recommend the usual OLS-based projection matrix; we use an alternative form proposed and justified by Saunders (2015).

The maximum likelihood estimates indicate that a negative relationship between the previous period's realization of costs and the current period and that the errors are positively skewed. The IIE parameter estimates indicate that there is no dynamic relationship between periods for the bank's costs. The more prominent positive skew for the IIE approach matches the upward bias correction observed in table 5.1.

The confidence intervals, constructed by the ICSI approach, fail to reject $\rho = 0$, which coincides with a static translog model. This is important, since the MLE approach indicates that the dynamic parameter is significant due to the downward bias of MLE. The shape parameter, λ , does not cover zero, which indicates that the normal distribution is not covered by the confidence set.

5.6 Final Remarks

We show that the IIE and ICSI approaches provide promising bias-correction and coverage for dynamic stochastic frontier models. Our analysis is seen as a first step to correct for known estimation bias in similar models. The proposed IIE approach uses MLE as the auxiliary estimator, similar to the fixed-effects estimator, while alternative estimators may improve efficiency of the ICSI method.

The time-varying technical (in)efficiency of the estimation is recoverable, in the manner described in Desli et al. (2003), but the proposed method allows for more elaborate error structure even in the presence of biased-MLE estimation. The unobserved heterogeneity in the proposed framework can be recovered recursively based on the unbiased parameter estimates. The IIE framework allows for non-Gaussian and Gaussian distribution assumptions on the unobserved heterogeneity.

6. CONCLUDING REMARKS

This thesis demonstrates again that dynamic panel data models are different from time series. The dual index leads to unique problems, like the incidental parameter bias, that require an alternative solution. The augmentation of the model as shown in chapter 3 is incompatible with time series and most vector autoregression models due to an excess of observations in time.

The augmentation is suitable for methods outside of the LSDV estimation and II estimation. The moment conditions for the bias-correction approaches, like Hahn and Kuersteiner (2002) and Bun and Carree (2005), are valid under the augmented model, which for Hahn and Kuersteiner (2002) allows for the approach to be extended to include regressors. The augmented model is neutral on the main set of GMM conditions that currently exist in the dynamic panel literature, which will extend the approach of Bun and Kleibergen (2013) to include regressors. Other worthy extensions of the approaches presented include unbalanced panels, co-integrated panels, relaxing the mean stationary assumption, and comparisons to a block-diagonal Mundlak device.

LIST OF REFERENCES

LIST OF REFERENCES

- Seung C Ahn and Peter Schmidt. Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1):5–27, 1995.
- Seung C Ahn and Robin C Sickles. Estimation of long-run inefficiency levels: a dynamic frontier approach. *Econometric Reviews*, 19(4):461–492, 2000.
- Theodore Wilbur Anderson and Cheng Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982.
- Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.
- Manuel Arellano and Olympia Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.
- Marie-Claude Beaulieu, Jean-Marie Dufour, and Lynda Khalaf. Multivariate tests of mean–variance efficiency with possibly non-gaussian errors. *Journal of Business & Economic Statistics*, 25(4), 2007.
- Marie-Claude Beaulieu, Jean-Marie Dufour, and Lynda Khalaf. Identification-robust estimation and testing of the zero-beta capm. *The Review of Economic Studies*, 80: 892–924, 2013.
- Marie-Claude Beaulieu, Jean-Marie Dufour, and Lynda Khalaf. Exact confidence sets and goodness-of-fit methods for stable distributions. *Journal of Econometrics*, 181(1):3–14, 2014.
- Alok Bhargava, Louisa Franzini, and Wiji Narendranathan. Serial correlation and the fixed effects model. *The Review of Economic Studies*, 49(4):533–549, 1982.
- Richard Blundell and Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998.
- Maurice JG Bun and Martin A Carree. Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics*, 23(2):200–210, 2005.
- Maurice JG Bun and Frank Kleibergen. Identification and inference in moments based analysis of linear dynamic panel data models. 2013.
- Maurice JG Bun and Vasilis Sarafidis. Dynamic panel data models. In B.H. Baltagi, editor, *The Oxford Handbook of Panel Data*. Oxford University Press, Oxford, 2015.
- Maurice JG Bun and Frank Windmeijer. The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal*, 13(1): 95–126, 2010.

Robert B Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254, 1977.

Robert B Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43, 1987.

Evangelia Desli, Subhash C Ray, and Subal C Kumbhakar. A dynamic stochastic frontier production model with time-varying efficiency. *Applied Economics Letters*, 10(10):623–626, 2003.

Jean-Marie Dufour. Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica: Journal of the Econometric Society*, pages 475–494, 1990.

Jean-Marie Dufour. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, 133(2):443–477, 2006.

Jean-Marie Dufour and Pascale Valéry. Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. *Journal of Econometrics*, 150(2):193–206, 2009.

Ragnar Frisch and Frederick V Waugh. Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, pages 387–401, 1933.

A Ronald Gallant and George Tauchen. Which moments to match? *Econometric Theory*, 12(04):657–681, 1996.

Silvia Gonçalves and Maximilien Kaffo. Bootstrap inference for linear dynamic panel data models with individual fixed effects. *forthcoming in Journal of Econometrics*, 2014.

Yuriy Gorodnichenko, Anna Mikusheva, and Serena Ng. Estimators for persistent and possibly nonstationary data with classical properties. *Econometric Theory*, 28(05):1003–1036, 2012.

Christian Gourieroux and Alain Monfort. Testing, encompassing, and simulating dynamic econometric models. *Econometric Theory*, 11(02):195–228, 1995.

Christian Gourieroux, Alain Monfort, and Eric Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118, 1993.

Christian Gouriéroux, Peter CB Phillips, and Jun Yu. Indirect inference for dynamic panel models. *Journal of Econometrics*, 157(1):68–77, 2010.

Zvi Griliches and M. D. Intriligator, editors. *Handbook of Econometrics*, volume 2. Elsevier, 1 edition, 1984.

Jinyong Hahn. How informative is the initial condition in the dynamic panel model with fixed effects? *Journal of Econometrics*, 93(2):309–326, 1999.

Jinyong Hahn and Guido Kuersteiner. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. *Econometrica*, 70(4):1639–1657, 2002.

Jinyong Hahn and Hyungsik Roger Moon. Reducing bias of MLE in a dynamic panel model. *Econometric Theory*, 22(03):499–512, 2006.

Chirok Han and Peter CB Phillips. GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory*, 26(01):119–151, 2010.

Chirok Han, Peter CB Phillips, and Donggyu Sul. X-differencing and dynamic panel model estimation. *Econometric Theory*, 30(1):201–251, 2014.

Bruce E Hansen. Autoregressive conditional density estimation. *International Economic Review*, pages 705–730, 1994.

Lynda Khalaf and Charles J. Saunders. Confidence intervals in autoregressive panels, invariance, and indirect inference. *Manuscript*, 2014.

Jan F Kiviet. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics*, 68(1):53–78, 1995.

Jan F Kiviet and IT van den Doel. Asymptotic consequences of neglected dynamics in individual effects models*. *Statistica Neerlandica*, 48(1):71–85, 1994.

Hugo Kruiniger. GMM estimation and inference in dynamic panel data models with persistent data. *Econometric Theory*, 25(05):1348–1391, 2009.

Subal C Kumbhakar and Efthymios G Tsionas. Measuring technical and allocative inefficiency in the translog cost system: a bayesian approach. *Journal of Econometrics*, 126(2):355–384, 2005.

Michael C Lovell. Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304):993–1010, 1963.

Xun Lu and Liangjun Su. Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Available at SSRN 2562174*, 2015.

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026, 1989.

Marcelo J. Moreira. A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics*, 37(6A):3660–3696, 12 2009.

Whitney K Newey and Kenneth D West. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787, 1987.

Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426, 1981.

Peter CB Phillips. On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195, 2014a.

Peter CB Phillips. Dynamic panel GMM with roots near unity. Technical report, Working Paper, Yale University, 2014b.

Peter CB Phillips and Donggyu Sul. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 137(1):162–188, 2007.

David Roodman. A note on the theme of too many instruments*. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158, 2009.

Charles J. Saunders. On autoregressive panels with covariates: Invariance, indirect inference, and confidence sets. *Manuscript*, 2015.

Anthony A Smith. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8(S1):S63–S84, 1993.

James H Stock and Mark W Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica: Journal of the Econometric Society*, pages 783–820, 1993.

James H Stock and Jonathan H Wright. GMM with weak identification. *Econometrica*, pages 1055–1096, 2000.

Lorenzo Trapani. Inference on the number of factors in factor models. *Manuscript*, 2015.

Efthymios G Tsionas. Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics*, 21(5):669–676, 2006.

Frank Windmeijer. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1):25–51, 2005.

APPENDICES

A. PROOFS FOR CHAPTER 2

For the proofs below we formally define the demeaned lagged dependent series as

$$g_{i,t-1} := g_{i,t-1}(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) = \left(\sum_{r=0}^{t-2} \epsilon_{i,t-1-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-2} \epsilon_{i,s-1-r} \phi^r \right) + \left(\sigma_\gamma \epsilon_{i,0} \phi^{t-1} - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \phi^{s-1} \right),$$

which is distinct from $g_{i,t}$ since it also back-shifts the summation for the empirical average.

A.1 Proof of Theorem 2.3.1

Proof [Panel-AR(1) Scale Invariance] Let $\tilde{\epsilon}_{i,t} = \kappa \epsilon_{i,t}$ and $\tilde{\epsilon}_{i,0} = \kappa \epsilon_{i,0}$, where κ is a positive constant, leading to the transformed series $\tilde{y}_{i,t}$. The transformed demeaned series are defined as,

$$\begin{aligned} \tilde{y}_{i,t}^* &= \tilde{y}_{i,t} - \tilde{y}_{i,\bullet} \\ &= g_i(\phi, \sigma_\gamma, \tilde{\epsilon}_{i,0}, \tilde{\epsilon}_i, t, T) = g_i(\phi, \sigma_\gamma, \kappa \epsilon_{i,0}, \kappa \epsilon_i, t, T) \\ &= \left(\sum_{r=0}^{t-1} \tilde{\epsilon}_{i,t-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \tilde{\epsilon}_{i,t-r} \phi^r \right) + \left(\sigma_\gamma \tilde{\epsilon}_{i,0} \phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \tilde{\epsilon}_{i,0} \phi^s \right) \\ &= \left(\sum_{r=0}^{t-1} \kappa \epsilon_{i,t-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \kappa \epsilon_{i,t-r} \phi^r \right) + \left(\sigma_\gamma \kappa \epsilon_{i,0} \phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \kappa \epsilon_{i,0} \phi^s \right) \\ &= \kappa \left[\left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r \right) + \left(\sigma_\gamma \epsilon_{i,0} \phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \phi^s \right) \right] \\ &= \kappa g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) \\ &= \kappa y_{i,t}^*, \end{aligned}$$

and likewise for the lagged dependent series. The series is linear in σ_ϵ , so κ is a common factor for all the terms in the series. The LSDV estimator of ϕ is

$$\begin{aligned}\tilde{\phi} &= (\kappa y_{-1}^{*'} \kappa y_{-1}^*)^{-1} \kappa y_{-1}^{*'} \kappa y^* \\ &= \frac{\kappa^2}{\kappa^2} (y_{-1}^{*'} y_{-1}^*)^{-1} y_{-1}^{*'} y^* \\ &= \hat{\phi}\end{aligned}$$

which is identical to Equation 2.5. Therefore under assumption 1 the LSDV estimator of ϕ is scale invariant. ■

Proof [Panel-AR(1) Location Invariance] Let $\tilde{\mu}_i$ represent a transformed set of random coefficients, where the transformation can be either a level shift, $\tilde{\mu}_i = \delta_i + \mu_i$, or relative shift, $\tilde{\mu}_i = \kappa_i \mu_i$, or both, $\tilde{\mu}_i = \delta_i + \kappa_i \mu_i$. The new demeaned series, $\tilde{y}_{i,t}$, follows as above leading to

$$\begin{aligned}\tilde{y}_{i,t}^* &= \tilde{y}_{i,t} - \tilde{y}_{i,\bullet} \\ &= \left[\tilde{\mu}_i + \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r \right) + \sigma_\gamma \epsilon_{i,0} \phi^t \right] - \left[\tilde{\mu}_i + \left(\frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r \right) + \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \phi^s \right] \\ &= \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r \right) + \left(\sigma_\gamma \epsilon_{i,0} \phi^t - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \phi^s \right) \\ &= g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) \\ &= y_{i,t}^*\end{aligned}$$

and likewise for the lagged dependent series. The demeaned series are independent of $\tilde{\mu}$ and equal to the demeaned series if $\delta_i \neq 0$ or $\kappa_i = 1$. Hence, the LSDV estimator of ϕ from the transformed data is identical to that of the untransformed data. Therefore under assumptions 1 the LSDV estimator of ϕ is location invariant. ■

A.2 Corollary Proofs

Proof [Proof of Corollary 1: Affine Transformation Invariance] The LSDV estimator is invariant to affine transformations of the data. Consider the transformation $\tilde{y}_{i,t} = (\kappa y_{i,t} + \delta_i)$. The transformed series can be defined as,

$$\begin{aligned}\tilde{y}_{i,t} &= \kappa\mu_i + \kappa \sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r + \kappa\sigma_\gamma \epsilon_{i,0} \phi^t + \delta_i \\ &= (\kappa\mu_i + \delta_i) + \kappa \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r + \sigma_\gamma \epsilon_{i,0} \phi^t \right).\end{aligned}$$

The empirical average is given by

$$\tilde{y}_{i,\bullet} = (\kappa\mu_i + \delta_i) + \kappa \frac{1}{T-1} \sum_{s=2}^T \left(\sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r + \sigma_\gamma \epsilon_{i,0} \phi^s \right),$$

which leads to the demeaned series,

$$\begin{aligned}\tilde{y}_{i,t}^* &= \tilde{y}_{i,t} - \tilde{y}_{i,\bullet} \\ &= \kappa \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} \phi^r + \sigma_\gamma \epsilon_{i,0} \phi^t \right) - \kappa \frac{1}{T-1} \sum_{s=2}^T \left(\sum_{r=0}^{s-1} \epsilon_{i,t-r} \phi^r + \sigma_\gamma \epsilon_{i,0} \phi^s \right) \\ &= \kappa g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T) \\ &= \kappa y_{i,t}^*,\end{aligned}$$

and likewise for the lagged dependent series. By scale invariance of the estimator, we find that $\tilde{\phi} = \hat{\phi}$, so the LSDV estimator is invariant to affine transformations. ■

Proof [Proof of Corollary 2: Invariance to σ_γ at Unity] The LSDV estimator is invariant to σ_γ for $\phi = 1$. Consider the demeaned series under unity,

$$\begin{aligned}
 y_{i,t}^*|_{\phi=1} &= g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T | \phi = 1) \\
 &= \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \right) + \left(\sigma_\gamma \epsilon_{i,0} - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \right) \\
 &= \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \right) + (\sigma_\gamma \epsilon_{i,0} - \sigma_\gamma \epsilon_{i,0}) \\
 &= \left(\sum_{r=0}^{t-1} \epsilon_{i,t-r} - \frac{1}{T-1} \sum_{s=2}^T \sum_{r=0}^{s-1} \epsilon_{i,t-r} \right),
 \end{aligned}$$

and likewise for the lagged dependent series. Therefore the LSDV estimator of ϕ is not a function of σ_γ , as such simulated series under the unit root null allow for any choice of σ_γ . By setting $\sigma_\gamma = 0$ when $\phi = 1$, we recover the conditional setting outlined in assumption 3. ■

B. PROOFS FOR CHAPTER 4

The demeaned lagged dependent series by recursive substitution is

$$y_{i,t-1}^* = a_{i,t-1}\beta + g_{i,t-1}, \text{ where} \quad (\text{B.1})$$

$$a_{i,t-1} = \left(\sum_{j=0}^{t-2} x'_{i,t-j-1} \phi^j - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-2} x'_{i,s-j-1} \phi^j \right) \text{ and} \quad (\text{B.2})$$

$$\begin{aligned} g_{i,t-1} = & \left(\sum_{j=0}^{t-2} \epsilon_{i,t-j-1} \phi^j - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-2} \epsilon_{i,s-j-1} \phi^j \right) \\ & + \left(\sigma_{\gamma} \epsilon_{i,0} \phi^{t-1} - \frac{1}{T-1} \sum_{s=2}^T \sigma_{\gamma} \epsilon_{i,0} \phi^{s-1} \right). \end{aligned} \quad (\text{B.3})$$

The matrix representation of the demeaned model is

$$y_{-1}^* = A_{-1}\beta + g_{-1}. \quad (\text{B.4})$$

I obtain the alternative representations of the lag dependent variables. The resulting equations will have the same W_X matrix, but the weights are shifted, so

$$\omega_{-1} = (\omega'_{1,-1}, \omega'_{2,-1}, \dots, \omega'_{T-1,-1})' \text{ and} \quad (\text{B.5})$$

$$\omega_{t,-1} = (\omega_{t,1,-1}, \omega_{t,2,-1}, \dots, \omega_{t,T,-1})' \text{ where} \quad (\text{B.6})$$

$$\omega_{t,r,-1} = \begin{cases} \left(\frac{1-\phi^{T-r}}{(T-1)(1-\phi)} + \phi^{t-r} \right) & \text{if } 1 \leq r \leq t \\ \left(\frac{1-\phi^{T-r}}{(T-1)(1-\phi)} \right) & \text{if } r > t \\ 0 & \text{if } r = T. \end{cases} \quad (\text{B.7})$$

The two complementary lag dependent equation forms are given by,

$$y_{-1}^* = A_{-1}\beta + g_{-1} \quad (\text{B.8})$$

$$y_{-1}^* = W_X W_{\omega,-1} \beta + g_{-1}, \quad (\text{B.9})$$

where $W_{\omega,-1} = [\iota_K \otimes \omega_{-1}]$.

B.1 Proof of Theorem 4.4.1

Proof [Augmented-LSDV Scale Invariance] Let $\kappa_1 > 0$ and $\kappa_2 > 0$ represent scale factors for ϵ and x respectively, leading to $\tilde{\epsilon}_{i,t} = \kappa_1 \epsilon_{i,t}$ and $\tilde{x}_{i,t} = \kappa_2 x_{i,t}$. As shown in Section 4.3, $\tilde{A}$ is linear in $\tilde{x}$ so κ_2 is a common factor,

$$\tilde{A} = \kappa_2 A,$$

so $M_X \tilde{A} = \kappa_2 M_X A = 0$. So any positive scalar multiplied to the exogenous regressors are partialled out with M_X . So the augmented-LSDV is invariant to scale transformations of the exogenous regressors.

Similarly, $\tilde{g} = \kappa_1 g$ and $\tilde{g}_{-1} = \kappa_1 g_{-1}$, so the scaled augmented-LSDV estimator of ϕ is

$$\begin{aligned} \tilde{\phi} &= (\tilde{g}'_{-1} M_X \tilde{g}_{-1})^{-1} \tilde{g}'_{-1} M_X \tilde{g} \\ \tilde{\phi} &= (\kappa_1^2 g'_{-1} M_X g_{-1})^{-1} \kappa_1^2 g'_{-1} M_X g \\ \tilde{\phi} &= (g'_{-1} M_X g_{-1})^{-1} g'_{-1} M_X g \\ \tilde{\phi} &= \hat{\phi}. \end{aligned}$$

Therefore for dynamic panel models with exogenous regressors, under assumptions 5, 4, 6 and $\beta \neq 0$, the augmented-LSDV estimator is scale invariant. ■

Proof [Augmented-LSDV Location Invariance] Let $\tilde{\mu}_i$ represent a transformed set of time-invariant error terms and $\tilde{x}_{i,0}$ represent a transformed set of initial exogenous

regressor observations, where the transformation can be either a level shift, $\tilde{\mu}_i = \delta_{i,1} + \mu_i$, $\tilde{x}_{i,0} = \delta_{i,2} + x_{i,0}$, or relative shift, $\tilde{\mu}_i = \kappa_{i,1}\mu_i$, $\tilde{x}_{i,0} = \kappa_{i,2}x_{i,0}$, or both, $\tilde{\mu}_i = \delta_{i,1} + \kappa_{i,1}\mu_i$, and $\tilde{x}_{i,0} = \delta_{i,2} + \kappa_{i,2}x_{i,0}$. The new demeaned series, $\tilde{y}_{i,t}$, follows as above leading to

$$\begin{aligned}
\tilde{y}_{i,t}^* &= \tilde{y}_{i,t} - \tilde{y}_{i,\bullet} \\
&= (\tilde{\mu}_i + \tilde{x}_{i,0}\beta) - (\tilde{\mu}_i + \tilde{x}_{i,0}\beta) + a_{i,t}\beta + g_{i,t} \\
&= a_{i,t}\beta + g_{i,t} \\
&= y_{i,t}^*,
\end{aligned}$$

and likewise for the lagged dependent series. The transformed demeaned series are independent of $\tilde{\mu}$ and $\tilde{x}_{i,0}$, and equal to the original demeaned series where $\delta_{i,1} = \delta_{i,2} = 0$, $\kappa_{i,1} = \kappa_{i,2} = 1$. Hence, the augmented-LSDV estimator of ϕ from the transformed data is identical to the untransformed data. Therefore under assumption 4 the LSDV estimator of ϕ is location invariant. ■

B.2 Corollary Proofs

Proof [Proof of Corollary 3: Affine Transformation Invariance] The augmented-LSDV estimator is invariant to affine transformations of the data. Consider the transformation $\tilde{y}_{i,t} = (\kappa y_{i,t} + \delta_i)$. The transformed series can be defined as,

$$\begin{aligned}
\tilde{y}_{i,t} &= \kappa(1 - \phi) (\mu_i + x'_{i,0}\beta) + \kappa\phi y_{i,t-1} + \kappa x'_{i,t}\beta + \kappa\epsilon_{i,t} + \delta_i \\
&= (\kappa(1 - \phi) (\mu_i + x'_{i,0}\beta) + \delta_i) + \kappa (\phi y_{i,t-1} + x'_{i,t}\beta + \epsilon_{i,t}).
\end{aligned}$$

The empirical average is given by

$$\tilde{y}_{i,\bullet} = (\kappa(1 - \phi) (\mu_i + x'_{i,0}\beta) + \delta_i) + \kappa \frac{1}{T-1} \sum_{s=2}^T (\phi y_{i,s-1} + x'_{i,s}\beta + \epsilon_{i,s}),$$

which leads to the demeaned series,

$$\begin{aligned}
 \tilde{y}_{i,t}^* &= \tilde{y}_{i,t} - \tilde{y}_{i,\bullet} \\
 &= \kappa a_{i,t} \beta + \kappa g_{i,t} \\
 &= \kappa y_{i,t}^*,
 \end{aligned}$$

and likewise for the lagged dependent series. By scale invariance of the estimator, we find that $\tilde{\phi} = \hat{\phi}$, so the augmented-LSDV estimator is invariant to affine transformations. ■

Proof [Proof of Corollary 4: Invariance to σ_γ at Unity] The augmented-LSDV estimator is invariant to σ_γ for $\phi = 1$. Consider the demeaned series under unity,

$$\begin{aligned}
 y_{i,t}^*|_{\phi=1} &= a_i(\phi, x_i, t, T|\phi=1)\beta + g_i(\phi, \sigma_\gamma, \epsilon_{i,0}, \epsilon_i, t, T|\phi=1) \\
 &= \left(\sum_{j=0}^{t-1} x'_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} x'_{i,s-j} \right) \beta + \left(\sum_{j=0}^{t-1} \epsilon_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} \epsilon_{i,s-j} \right) \\
 &\quad + \left(\sigma_\gamma \epsilon_{i,0} - \frac{1}{T-1} \sum_{s=2}^T \sigma_\gamma \epsilon_{i,0} \right) \\
 &= \left(\sum_{j=0}^{t-1} x'_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} x'_{i,s-j} \right) \beta + \left(\sum_{j=0}^{t-1} \epsilon_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} \epsilon_{i,s-j} \right) \\
 &\quad + (\sigma_\gamma \epsilon_{i,0} - \sigma_\gamma \epsilon_{i,0}) \\
 &= \left(\sum_{j=0}^{t-1} x'_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} x'_{i,s-j} \right) \beta + \left(\sum_{j=0}^{t-1} \epsilon_{i,t-j} - \frac{1}{T-1} \sum_{s=2}^T \sum_{j=0}^{s-1} \epsilon_{i,s-j} \right),
 \end{aligned}$$

and likewise for the lagged dependent series. Therefore the augmented-LSDV estimator of ϕ is not a function of σ_γ , as such simulated series under the unit root null allow for any choice of σ_γ . By setting $\sigma_\gamma = 0$ when $\phi = 1$, we recover the conditional setting outlined in assumption 6. ■