

# Estimating Panel Models with Internal and External Habit Formation

George M. Korniotis\*

Board of Governors of the Federal Reserve System

February 20, 2007

A new bias-corrected estimator is developed to test for habit formation. The estimator is applicable to any dynamic panel model with fixed and spatial effects as well as endogenous control variables. The estimator is asymptotically unbiased and normally distributed. Moreover, simulation results demonstrate that it has low finite-sample bias and low root mean squared error. The estimator is ideal for testing for habit formation in consumption growth data because it is designed for models that include a time-lagged dependent variable, measuring internal-habit formation, and a spatially lagged dependent variable, measuring external-habit formation. Applying the estimator to annual consumption data for the continental U.S. states shows that state consumption growth is not significantly affected by its own (lagged) consumption growth. However, it is affected by the consumption of other states. In particular, the closer two states are, the more they affect each other.

**KEY WORDS:** Dynamic Panel Models, Bias Correction, Spatial Effects, Habit Formation, U.S. State Consumption

---

\*The paper has benefited from comments by Elena Andreou, Don Andrews, George Constantinides, Tom Cosimano, Gregg Forte, Rubén Hernández-Murillo, Erik Hjalmarsson, Ruth Judson, Guido Kuersteiner, Alok Kumar, Yoonseok Lee, Michael McCracken, Peter Phillips, Aris Spanos, Mary Beth Walker, Jonathan Wright, and the participants of seminars at the Department of Economics at Yale University, the 2005 Workshop on Spatial Econometrics at Kiel, Germany, the 2006 Meeting of the Econometric Society, Boston, and the 2007 meetings of the Southern Economic Association, New Orleans. I also would like to thank the editor, Torber Andersen, the anonymous associate editor and referee for their feedback. Finally, I would like to thank Ruben Hernandez-Murillo for sharing the state population centroid data, which he uses in his 2003 paper. Any remaining errors are mine. Also, the analysis and conclusions set forth are those of the author and not necessarily those of the Board of Governors or its staff. Please address correspondence to George M. Korniotis, Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue NW, Washington, DC 20551. E-mail: George.M.Korniotis@frb.gov.

## 1. INTRODUCTION

This paper introduces a new bias-corrected estimator for dynamic panel models with fixed and spatial effects to test for habit formation at the aggregate level. Testing for habit formation is important because consumption-based asset pricing models with habits are one of the leading approaches in explaining asset pricing phenomena (Constantinides 2002). In these models, the welfare of the consumer depends on the difference between actual consumption and a habit level of consumption. The habit is a time-varying subsistence level, and it typically takes one of two forms. For models in which the determinants of the habit are internal to the consumer (internal habit), the consumers' consumption habits are influenced by *own* past consumption (Constantinides 1990; Ferson and Constantinides 1991; Ferson and Harvey 1992; Fuhrer 2000; Dynan 2000). For models in which the determinants of the habit are external to the consumer (external habit), consumers' consumption habits are influenced by the consumption decisions of *other* consumers (Abel 1990; Gali 1994; Campbell and Cochrane 1999; Chan and Kogan 2002; Wachter 2006). Even though internal and external habits are widely used, there are only a couple of studies estimating models that include both of them (Grishchenko 2004; Ravina 2005).

Since little is known about the relative importance of internal and external habits, I test for their significance by estimating Euler equations with data for the 48 continental U.S. states. As in Ferson and Constantinides (1991), the measure of the internal component of the habit is past own consumption. As in Abel (1990), the measure of the external component of the habit is a weighted average of past consumption growth rates of other cross-sectional units; this type of external habit is called "catching up with the Joneses." The external habit of a state is a spatially lagged dependent variable because it is defined by consumption growth rates of other states spatially located in the same national economy. The Euler equation of the economic model with both types of habit formation produces a dynamic panel regression in which the dependent variable is the consumption growth rate of each U.S. state. The regression includes fixed effects to allow for heterogeneous rates of consumption growth across the U.S. states. The regression involves a time-lagged dependent variable (the internal-habit measure) and a spatially lagged dependent variable (the external-habit measure). The regression also contains the contemporaneous return

of the one-month Treasury bill (the endogenous regressor).

The existing literature provides little guidance on estimating panel regression models with habit variables and endogenous regressors. I therefore introduce a new estimator, which extends the work of Hahn and Kuersteiner (2002) on estimating dynamic panel models with the least-squares dummy variable (LSDV) estimator. My extension, which allows for spatial effects and endogenous control variables, is a hybrid of the LSDV and the instrumental variable estimator of Anderson and Hsiao (1982). Like Hahn and Kuersteiner (2002), I rescale (de-mean) the data to eliminate the fixed effects from the estimation, and, like Anderson and Hsiao (1982), I instrument the endogenous control variables. I show that the hybrid estimator is asymptotically biased. I therefore use its asymptotic bias to define a bias-corrected estimator that is asymptotically unbiased and asymptotically normal. In practice, the bias-corrected estimator might be preferred to a pure instrumental variable (IV) estimator because it only instruments the explanatory variables whereas an IV estimator has to also instrument the time-lagged and spatially lagged dependent variables. Consequently, the IV might be more exposed to weak instrument biases compared to the bias-corrected estimator (Hahn, Hausman, and Kuersteiner (2002), Hahn and Hausman (2003)).

Moreover, using Monte Carlo simulations, I find that the bias-corrected estimator compares favorably to a pure IV-type estimator. First, I find that the bias-corrected estimator has low finite-sample bias and low standard deviation. It also has a smaller root mean squared error than the pure IV estimator of Anderson and Hsiao (1982). Second, I find that in the presence of measurement error, the finite-sample bias of the bias-corrected estimator increases. The estimator however, maintains a low root mean squared error, and again has a lower standard deviation than the pure IV estimator.

I apply the bias-corrected estimator to test for habit formation. Like most studies in the consumption-based asset pricing literature, I test for habit formation at the aggregate level. However, instead of using national data, I use annual data for the 48 continental U.S. states for the period 1966-98. State-level data have important advantages when testing for habit formation. Specifically, even though state-level data are aggregate data, they exhibit considerable cross-

sectional variation. Thus, for each state, I can define measures of external-habit formation that are *independent* of internal-habit measures. For instance, if I am considering New Jersey, an external-habit measure can be the average of past consumption choices of states other than New Jersey, and a measure for the internal component of the habit is the past consumption level of New Jersey itself. This clear distinction between the two habit measures, which is lost when using national data, provides a powerful way to determine the type of habit best supported by the data.

In the empirical application, I measure internal-habit formation by own past consumption. I also consider five external-habit measures that exclude own consumption. The first one is based on the past consumption of neighboring states. The second measure is inversely related to the geographic distance between the U.S. states (Hernández-Murillo 2003). In this case, the further apart two states are, the less they influence each other. As in Case (1991), Topa (2001), and Ravina (2005), these two habit measures test whether the external component of the habit is determined by geographic proximity. I also consider potential determinants of external-habit formation, other than proximity. In particular, the third measure recognizes that consumption trends might originate from urban centers and thus refines the distance-based habit using the percentage of a state's population living in urban areas. To further examine the effect of urban centers, the fourth measure deletes from the third measure all the states with a low degree of urbanization. Finally, in the fifth measure, the external habit is given by the average consumption of all states excluding the state in question. The fifth measure is almost identical to past U.S. consumption, which is a popular choice in many asset pricing models (Abel 1990; Campbell and Cochrane 1999; Wachter 2006).

The estimation finds supporting evidence for external-habit formation and provides only weak evidence of internal-habit formation. In particular, the results for the five external-habit measures indicate that the closer two states are, the more they affect each other, a result consistent with existing work (Case 1991; Ravina 2005). I also find suggestive evidence that states with population that predominantly lives in urban centers affect the consumption of other states the most.

Beyond testing for habit formation, the bias-corrected estimator is of practical relevance to a wide range of studies, such as country comparisons (Islam 1995; Lee, Pesaran, and Smith 1998) and microeconomic studies using synthetic cohort data (Mason and Fienberg 1985; Deaton 1997). Therefore, in section 2, I present the panel model in general terms. In section 3, I define the bias-corrected estimator and its asymptotic properties. In section 4, I collect all the results related to testing for habit formation at the U.S. state level, and in section 5, I present the results of a simulation exercise. The set-up of the simulation is based on the empirical results in section 4. In section 6, I conclude the discussion.

## 2. ECONOMETRIC MODEL

This paper introduces a procedure for estimating linear models that includes fixed effects, a time-lagged dependent variable, and a spatially lagged dependent variable:

$$Y_{it} = \pi_1 Y_{i,t-1} + \rho_1 \sum_{j=1}^N w_{ij} Y_{j,t-1} + X_{it} \lambda + c_i + \eta_{it}, \quad i = [1, \dots, N], \quad t = [1, \dots, T], \quad (1)$$

where  $Y_{it}$  is the dependent variable for cross-sectional unit  $i$  at time  $t$  (Section A.4 in the appendix provides a generalization of the model with  $m_1$  time-lagged dependent variables, and  $m_2$  spatially lagged dependent variables). The timing convention is that the available data for estimation is from  $t = 0$  to  $t = T$ . Like Hahn and Kuersteiner (2002), the observation of  $Y$  at  $t = 0$ ,  $Y_0$ , represents the initial condition of  $Y$ , and is taken to be nonstochastic and known to the econometrician. The random variable  $\eta_{it}$  is an IID error term with zero mean and variance  $\sigma_\eta^2$ . The constant  $c_i$  is the fixed effect of cross-sectional unit  $i$  that absorbs time-invariant characteristics of  $i$  influencing the dependent variable  $Y$ . The vector  $X_{it}$  is a  $1 \times K$  vector of endogenous control variables (see Condition 3 below for additional assumptions on  $X_{it}$ ). The  $K \times 1$  vector  $\lambda$  includes the parameters related to  $X_{it}$ . The set of control variables can include contemporaneous and time-lagged values of  $X$ .

The weight  $w_{ij}$  measures the importance of  $Y_{j,t-1}$  on  $Y_{it}$ . The weights are observed quantities, which are known to the econometrician, and they are therefore exogenous. Because the spatial

lag,  $\sum_{j=1}^N w_{ij} Y_{j,t-1}$ , is a weighted average of past consumption choices of other cross-sectional units, it is the measure of the catching-up habit. The weights  $w_{ij}$  are organized in the  $N \times N$  matrix  $W$  called the spatial matrix. The structure of  $W$  is explained in Condition 2 below.

The presence of the fixed effects allows for non-parametric estimation of the time-invariant differences between the cross-sectional units. I choose the fixed effects model, over the random effects one, because it is more appropriate for models with habit formation. The economic model in Section 4 reveals that the fixed effects are related to state-specific discount rates. Since the state-specific discount rates depend on state characteristics (e.g. demographics), I cannot treat them as independent realization of a random shock, which is the assumption made in random effects models (Hausman (1978), Greene (2003), and Hsiao (2003)). For similar reasons, Asdrubali, Sorensen, and Yosha (1996), and Ostergaard, Sorensen, and Yosha (2002) use the fixed effects specification when estimating panel models with state-level consumption data.

The econometric model (1) is different from the model in Hahn and Kuersteiner (2002) in two ways: It allows for endogenous control variables, and it includes the spatially lagged dependent variable  $\sum_{j=1}^N w_{ij} Y_{j,t-1}$ . In particular, Hahn and Kuersteiner (2002) propose a bias-corrected estimator for dynamic panel models with only fixed effects and no endogenous explanatory variables:  $Y_{it} = \pi_1 Y_{i,t-1} + X_{it} \lambda + c_i + \eta_{it}$ .

The econometric model (1) operates under three conditions: condition 1 includes the assumption about the regression error term, the initial conditions of  $Y$ , the fixed effects  $c$ , and the asymptotic sequences. Condition 2 deals with the spatial matrix  $W$ , and condition 3 presents the assumptions for the control variables  $X$ .

*Condition 1.* 1(1): The error term  $\eta_{it}$  is IID across  $N$  and  $T$ . All the moments of  $\eta_{it}$  exist and  $\mathbf{E} |\eta_{it}|^{2+\zeta} < \infty$  for some  $\zeta > 0$  and all  $i$  and  $t$ . 1(2): The limit of  $N/T$  exists (as  $N, T \rightarrow \infty$ ) and it is bounded between 0 and  $\infty$ , i.e.  $N, T \rightarrow \infty$  such that  $\lim(N/T) = k$ ,  $0 < k < \infty$ . 1(3):  $|\rho_1| + |\pi_1| < 1$ . 1(4):  $\max_i |Y_{i0}|^2 = O(\sqrt{N})$ . 1(5):  $\max_i |c_i|^\zeta = O(1)$ ,  $\zeta > 0$ .

*Remark 1:* Condition 1(1) does not allow any cross-sectional or time correlations in the error term. Because the scope of model (1) is to *estimate* any time correlations (through time-lagged

dependent variables) and any spatial correlations (through spatially lagged dependent variables), it is convenient to purge such effects from the error term. Among others, panel models with cross-sectional dependence in the error term are investigated in Conley (1999) and Phillips and Sul (2003).

*Remark 2:* Condition 1(2) assumes that  $T$  and  $N$  grow at a finite rate, which is in line with the asymptotic analysis in Hahn and Kuersteiner (2002). Allowing both  $T$  and  $N$  to go to infinity is a useful analytical device because letting  $T$  approach infinity identifies the fixed effects which vary across  $i$ , and letting  $N$  approach infinite identifies the parameter on the spatially lagged dependent variable (Anselin 2001).

*Remark 3:* Condition 1(3) is a stationarity assumption. Conditions 1(4) and 1(5) control the behavior of the initial conditions  $Y_0$  and the fixed effects  $c$  respectively.

*Condition 2: Spatial Matrix.* The matrix  $W$  is an  $N \times N$  matrix with  $w_{ij}$  being the element on its  $i^{th}$  row and  $j^{th}$  column. 2(1):  $W$  is a real non-negative matrix, that is  $w_{ij} \geq 0$  for all  $i$  and  $j$ . 2(2): All the diagonal elements of  $W$  are zero, that is  $w_{ii} = 0$  for all  $i$ . 2(3): The rows of  $W$  sum to one, that is  $\sum_{j=1}^N w_{ij} = 1$  for all  $i$ . 2(4) The maximum column sum of  $W$  is  $k_c$ , and  $k_c$  is finite and strictly less than  $k_0 = [(|\pi_1| + |\rho_1|)^{-1} - |\pi_1|] / |\rho_1|$ . 2(5): As  $N \rightarrow \infty$ ,  $W$  maintains the above properties.

*Remark 4:* Condition 2(1) holds for all spatial matrices. If  $w_{ij} > 0$ , then unit  $j$  affects unit  $i$ . If  $w_{ij} = 0$ , then unit  $j$  does not affect  $i$ . Condition 2(2) is a normalization, which implies that  $i$  does not affect itself. Condition 2(3) implies that spatial variables, like  $\sum_{j=1}^N w_{ij} Y_{j,t-1}$ , are weighted averages.

*Remark 5:* Conditions 2(3) and 2(4) restrict the degree of cross sectional correlation between the units. Note that in virtually all large sample theory one has to restrict the degree of permissible correlations. For example, see Potscher and Prucha (1997, chapters 5 and 6), and Kapoor, Kelejian, and Prucha (2004, assumption 4).

*Remark 6:* The bound on  $k_c$ ,  $k_0$ , is very weak and in practice it is very unlikely that it will be binding. For example, in the empirical application in Section 4, across the various spatial

matrices I consider, the largest value for  $k_c$  is 2.49, and the smallest value for  $k_0$  is 32. The specific value of  $k_0$  is chosen for technical reasons explained in the Appendix. See the analysis of Results 1 and 2 in Section A.1 of the Appendix.

*Remark 7:* Condition 2(5) restricts the asymptotic theory on the set of spatial matrices that satisfy all the aforementioned conditions.

*Remark 8:* Under Condition 2, the matrix  $W$  does not have to be symmetric. In fact, because the rows of  $W$  sum to one, it is unlikely that  $w_{ij}$  will equal to  $w_{ji}$ . Moreover, Condition 2 does not preclude cases where  $w_{ij} > 0$  and  $w_{ji} = 0$ . In this case,  $j$  is the leader and it affects, but it is not affected by,  $i$  the follower.

*Condition 3: Endogenous Control Variables.* The control variables  $X$  are correlated with the regression error term  $\eta$ :  $\mathbf{E}(X_{k,it}\eta_{js}) = \sigma_{k,x\eta}$  for  $i = j$ ,  $t = s$ , and  $\mathbf{E}(X_{k,it}\eta_{js}) = 0$  otherwise. The probability limit of the matrix  $(1/NT) \sum_{i=1}^N \sum_{t=1}^T (X_{it}^d)' (X_{it}^d)$  is finite and nonsingular ( $X_{it}^d = X_{it} - \bar{X}_i$ ,  $\bar{X}_i = T^{-1} \sum_{t=1}^T X_{it}$ ). All the moments of  $X_{it}$  exist and the expectations  $\sum_{k=1}^K \mathbf{E} |X_{k,it}|^{2+\zeta}$  and  $\sum_{k=1}^K \mathbf{E} |X_{k,it}\eta_{it}|^{2+\zeta}$  are bounded for some  $\zeta > 0$  and all  $i$  and  $t$  ( $K$  is the total number of control variables,  $X_{k,t}$  is an  $N \times 1$  vector with the observations of the  $k^{th}$  control at time period  $t$ ).

*Remark 9:* The assumption of endogenous control variables is not standard in the spatial econometrics literature. For example, both Kapoor, Kelejian, and Prucha (2004) and Baltagi, Song, Jung, and Koh (2004) assume that the control variables are exogenous. However, in testing for habit formation, most control variables are endogenous. Such a variable is the interest rate on Treasury bills, which is affected by the policies of the Federal Reserve. These policies are themselves determined by economy-wide shocks that also influence state-level consumption growth, which is the dependent variable in the regression of the habit model.

### 3. BIAS-CORRECTED ESTIMATOR

This section develops the bias-corrected estimator, which is based on the least squares dummy



variable (LSDV) estimator. The typical LSDV estimate of  $\varphi$  is:

$$\hat{\varphi}_{LSDV} = \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{X}_{i,t-1}^d \right)' \tilde{X}_{i,t-1}^d \right]^{-1} \times \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{X}_{i,t-1}^d \right)' Y_{it}^d \right], \quad (2)$$

where the superscript  $d$  denotes data that have been rescaled (de-meaned) to have zero mean. The matrix  $\tilde{X}_{i,t-1}$  is a  $1 \times (K+2)$  data matrix equal to  $(Y_{i,t-1}, W_i Y_{t-1}, X_{it})$ ,  $X_{it}$  is a  $1 \times K$  vector,  $W_i$  is the  $i^{th}$  row of the spatial matrix  $W$ , and  $Y_{t-1} = [Y_{1,t-1}, \dots, Y_{N,t-1}]'$ .

Unfortunately, the LSDV is a biased estimator of  $\varphi$ . Its bias originates from the presence of fixed effects, which gives rise to the incidental parameter bias, and from the presence of endogenous control variables. To eliminate the incidental parameter bias, I follow the bias-correction method of Hahn and Kuersteiner (2002). Because Hahn and Kuersteiner (2002) do not allow for endogenous regressors, I extend their approach to accommodate for endogeneity. This is done in two steps. First, I define a hybrid estimator, a modification of LSDV, that does not suffer from endogeneity issues. Second, because the hybrid estimator only suffers from the incidental parameter bias, I can apply the bias-correction approach of Hahn and Kuersteiner (2002) to obtain an asymptotically unbiased estimator of  $\varphi$ .

### 3.1 Hybrid Estimator

I define a hybrid estimator that modifies the LSDV by instrumenting the control variables. The hybrid estimator is given by

$$\hat{\varphi}_b = \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \tilde{X}_{i,t-1}^d \right]^{-1} \times \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' Y_{it}^d \right],$$

where  $\tilde{Z}_{i,t-1}$  is  $(Y_{i,t-1}^d, W_i Y_{t-1}^d, Z_{i,t-1}^d)$ . The only difference between  $\hat{\varphi}_{LSDV}$  and  $\hat{\varphi}_b$  is that  $X_{it}$  is replaced (instrumented) with  $Z_{i,t-1}$ , which is a  $1 \times K$  vector of instruments for the endogenous variables  $X_{it}$ .

The  $\hat{\varphi}_b$  is a hybrid between the LSDV and the instrumental variables estimator by Anderson and Hsiao (1982). As in the LSDV, the hybrid estimator eliminates the fixed effects by de-meaning the data. As in Anderson and Hsiao (1982), who proposed an instrument variable

estimator for dynamic panel models with only fixed effects, it accounts for endogeneity by instrumenting the endogenous control variables. In particular, the instruments for the regressors dated at time  $t$  can be any variable dated at time  $t - \tau$ ,  $\tau > 0$ . The instruments for regressors dated at  $t - \tau$ ,  $\tau > 0$ , can be the regressors themselves. The instruments  $Z$  behave according to condition 4.

*Condition 4: Instruments.* The instruments  $Z$  are contemporaneously correlated with the regression error term  $\eta$ :  $\mathbf{E}(Z_{k,it}\eta_{js}) = \sigma_{kz\eta}$  for  $i = j$ ,  $t = s$ , and  $\mathbf{E}(Z_{k,it}\eta_{js}) = 0$  otherwise. The probability limit of the matrix  $(1/NT) \sum_{i=1}^N \sum_{t=1}^T (Z_{it}^d)' (X_{it}^d)$  is finite and nonsingular. All the moments of  $Z_{it}$  exist and for all  $i$ ,  $t$ , and  $\zeta > 0$ , the expectations  $\sum_{k=1}^K \mathbf{E}|Z_{k,it}|^{2+\zeta}$  and  $\sum_{k=1}^K \mathbf{E}|Z_{k,it}\eta_{it}|^{2+\zeta}$  are bounded.

*Remark 10:* In condition 4, I assume that the contemporaneous values of the instruments,  $Z_t$ , are correlated with the error term  $\eta_t$ . I make this assumption because in practice it is difficult to find instruments, which are orthogonal to the error term. However, time-lagged values of  $Z_t$  are taken to be independent of  $\eta_t$  and they can serve as valid instruments. The orthogonality between  $Z_{t-s}$  and  $\eta_t$ ,  $s > 0$ , is consistent with rational expectation models where error terms like  $\eta_t$  are interpreted as forecast errors, which have to be orthogonal to any past information (see the discussion in Section 4.2).

Having defined the instruments for the endogenous regressors, I follow Hahn and Kuersteiner (2002) and I show that the  $\hat{\varphi}_b$  has a finite asymptotic bias. I also use arguments similar to Driscoll and Kraay (1998) to establish that the uncorrected hybrid estimator converges to a normal distribution.

*Theorem 1.* Under conditions 1 to 5,  $\sqrt{NT}(\hat{\varphi}_b - \varphi) \rightarrow \mathcal{N}(B, Q^{-1}VQ^{-1})$ , where  $B$  is a vector of bias terms, and  $Q = \text{plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \tilde{X}_{i,t-1}^d$ ,  $V = \text{plim} \frac{1}{NT} [\bar{S} - \mathbf{E}\bar{S}] [\bar{S} - \mathbf{E}\bar{S}]'$ ,  $\bar{S} = \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \eta_t^d$ . *Proof:* See appendix A.3. Note that condition 5 is in the appendix.

The  $\hat{\varphi}_b$  estimator is consistent but has a limiting distribution that is not centered at zero.

The asymptotic bias vector  $B$  is the limit of  $B_{NT}$ :

$$B_{NT} = \left[ -\frac{\sigma_\eta^2 + \boldsymbol{\sigma}_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr}\{\Pi\}, \quad -\frac{\sigma_\eta^2 + \boldsymbol{\sigma}_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr}\{W\Pi\}, \quad -\sqrt{\frac{N}{T}} \frac{T-1}{T} \boldsymbol{\sigma}_{z\eta} \right]', \quad (3)$$

where  $\mathbf{tr}$  is the trace operator. The  $1 \times K$  vector  $\boldsymbol{\sigma}_{x\eta} = (\sigma_{1,x\eta}, \dots, \sigma_{K,x\eta})$  contains the expectations  $\sigma_{k,x\eta} = \mathbf{E}(X_{k,it}\eta_{it})$ ,  $k = \{1, \dots, K\}$ . The  $1 \times K$  vector  $\boldsymbol{\sigma}_{z\eta} = (\sigma_{1,z\eta}, \dots, \sigma_{K,z\eta})$  contains the expectations  $\sigma_{k,z\eta} = \mathbf{E}(Z_{k,it}\eta_{it})$ ,  $k = \{1, \dots, K\}$  ( $Z_k$  is the instrument for the  $k^{th}$  control variable  $X_k$ ). Also, the matrix  $\Pi$  is  $(I - \Psi)^{-1}$ ,  $\Psi = (\pi_1 I + \rho_1 W)$ .

### 3.2 Bias-Correction

Theorem 1 shows that the asymptotic bias of the hybrid estimator is bounded. I therefore use the bias vector (3), and I define the subsequent bias-corrected estimator:

$$\hat{\varphi}_c = \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \tilde{X}_{i,t-1}^d \right]^{-1} \times \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' Y_{it}^d - \frac{\hat{B}_{NT}}{\sqrt{NT}} \right], \quad (4)$$

where  $\hat{B}_{NT}$  is the value of  $B_{NT}$  under a consistent estimator of  $\varphi$ ,  $\hat{\varphi}_0$ . Unlike the  $\hat{\varphi}_b$  estimator, the bias-corrected estimator  $\hat{\varphi}_c$  is consistent, and its distribution is centered at zero:

*Theorem 2.* Under conditions 1 to 5,  $\sqrt{NT}(\hat{\varphi}_c - \varphi) \rightarrow \mathcal{N}(0, Q^{-1}VQ^{-1})$ , where  $Q$  and  $V$  are as in theorem 1. *Proof:* See appendix A.3. Note that condition 5 is in the appendix.

### 3.3 Implementation

To implement the bias-corrected estimator  $\hat{\varphi}_c$ , one needs a consistent estimator of  $\varphi$ ,  $\hat{\varphi}_0$ , to calculate the bias vector  $B_{NT}$ . It is important to choose a  $\hat{\varphi}_0$ , which has good finite sample properties, to ensure that the finite sample performance of the corrected estimator  $\hat{\varphi}_c$  is not hindered by any finite sample biases in  $\hat{\varphi}_0$ . Moreover, one needs a consistent estimate of  $V$  to calculate the standard errors of  $\hat{\varphi}_c$ .

First, to obtain  $\hat{\varphi}_0$ , I modify the Anderson and Hsiao (1982) (AH) estimator to accommodate for spatial effects. Specifically, I take first differences of regression (1) to factor out the fixed effects ( $\Delta Y_t = \Delta \tilde{X}_{t-1}\varphi + \Delta \eta_t$ ), and I then estimate the first-difference model using instrumental

variables. The Anderson and Hsiao (1982) estimator for  $\pi_1$ ,  $\rho_1$ , and  $\lambda$  is:

$$\hat{\varphi}_{AH} = \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-2} \right)' \Delta \tilde{X}_{i,t-1} \right]^{-1} \times \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-2} \right)' \Delta Y_{it} \right], \quad (5)$$

where  $\tilde{Z}_{i,t-2} (= (Y_{i,t-2}, W_i Y_{t-2}, Z_{i,t-2}))$  includes the instruments for  $\Delta \tilde{X}_{i,t-1} (= (\Delta Y_{i,t-1}, W_i \Delta Y_{i,t-1}, \Delta X_t))$ . I use the residuals  $\hat{\eta}_t (= Y_t^d - \tilde{X}_t^d \hat{\varphi}_{AH})$  to estimate  $\sigma_\eta^2$ . To obtain estimates of  $\sigma_{x\eta}$  and  $\sigma_{z\eta}$ , I respectively use the sample averages of the cross-products of  $X_t$  and  $Z_t$  with  $\Delta \hat{\eta}_t$ , where  $\Delta \hat{\eta}_t$  is the residual from the AH estimation ( $\Delta \hat{\eta}_t = \Delta Y_t - \Delta \tilde{X}_{t-1} \hat{\varphi}_{AH}$ ). Kiviet (1995) uses a similar approach to calculate the bias-correction terms for the LSDV estimator in dynamic panel models with only fixed effect.

*Remark 11:* An alternative way to calculate  $\sigma_{x\eta}$  and  $\sigma_{z\eta}$  is with cross-products with  $\hat{\eta}_t$ . In unreported results however, I find that this alternative approach has almost no impact on the empirical results reported in Sections 4.3 and 4.4.

*Remark 12:* An alternative choice for  $\hat{\varphi}_0$  is the biased LSDV  $\hat{\varphi}_b$ , which is also consistent according to Theorem 1. In unreported simulation results, available upon request, I find that the percentage finite sample bias of  $\hat{\varphi}_b$  is higher than the finite sample bias of the  $\hat{\varphi}_{AH}$ . Therefore, in the estimations presented in Section 4 I use  $\hat{\varphi}_{AH}$  for the initial estimator  $\hat{\varphi}_0$ .

Second, to calculate the standard errors of  $\hat{\varphi}_c$ , I obtain a consistent estimate of  $V$  following Driscoll and Kraay (1998). Driscoll and Kraay (1998) extend Andrew's (1991) heteroscedasticity-consistent and autocorrelation-consistent (HAC) estimator of the covariance matrix of the score function of panel models. The Driscoll and Kraay (1998) estimator is implemented in three steps. First, at each point in time one calculates the cross-sectional averages  $h_{N,t}$ ,

$$h_{N,t} = \frac{1}{N} \sum_{i=1}^N \left( \tilde{Z}_{i,t-1}^d \right)' \left( Y_{it}^d - \tilde{X}_{i,t-1}^d \hat{\varphi}_c \right).$$

Next, using the time-series of these cross-sectional averages,  $h_{N,t}$ , one applies Andrew's (1991) HAC methodology to obtain the variance-covariance matrix of  $h_{N,t}$ ,  $V_h$ . Finally, one uses the

sandwich formula,

$$\left[ \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \left( \tilde{Z}_{i,t-1}^d \right)' \tilde{X}_{i,t-1}^d \right]^{-1} \frac{V_h}{T} \left[ \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \left( \tilde{X}_{i,t-1}^d \right)' \tilde{Z}_{i,t-1}^d \right]^{-1},$$

to calculate the variance-covariance matrix of the bias corrected estimator  $\hat{\varphi}_c$ .

#### 4. HABIT FORMATION AT THE U.S. STATE LEVEL

A natural application of the bias-corrected estimator is estimating models with habit formation because the time-lagged dependent variable is a measure of internal-habit formation, and the spatially lagged dependent variable is a measure of external-habit formation. Testing for habit formation is important because habit formation models are one of the leading consumption-based asset-pricing model that can explain the behavior of U.S. asset prices. Apart from explaining asset pricing phenomena, the concept of habit formation has been used to understand the persistence of aggregate output (Boldrin, Christiano, and Fisher 2001), the relationship between savings and growth (Carroll, Overland, and Weil 2000), and the reaction of aggregate consumption to monetary shocks (Fuhrer 2000).

Having set up the necessary econometric theory in sections 2 and 3, I now construct a habit formation model to explain U.S. state consumption growth. In the model, the economy is populated by 48 heterogeneous state consumers, one for each state. Next, I define the preferences of the state consumers and derive the econometric model.

##### 4.1 Preferences

The utility function of the representative consumer of state  $i$  is defined in terms of the difference between actual consumption and the habit level of consumption:

$$U_{it} = \frac{e^{-\beta_i}}{1-\gamma} [C_{it} - \rho_1 W_i C_{t-1} - \pi_1 C_{i,t-1}]^{1-\gamma}, \quad (6)$$

where  $\gamma$  is the curvature parameter, and  $C$  is consumption. Related to  $\gamma$  is the coefficient of relative risk aversion, which is given by  $\gamma/S_{it}$ ,  $S_{it} = 1 - (\rho_1 W_i C_{t-1} + \pi_1 C_{i,t-1})/C_{it}$  (See Campbell

and Cochrane, 1999). The term  $\exp(-\beta_i)$  is the factor of time preference, and  $\beta_i$  is the discount rate ( $\beta_i < 1$ ), which differs across the state consumers. Allowing the degree of time preference to vary across the state consumers is in line with microeconomic studies like Zeldes (1989) and Attanasio and Weber (1995).

The habit level of consumption includes an internal and an external term. Following Ferson and Constantinides (1991), the time lag of consumption,  $C_{i,t-1}$ , is the measure for internal-habit formation. The presence of the internal consumption habit in the utility function (6) implies that, all else equal, consumers' want to consume more compared to their past consumption. Therefore, they dislike large changes in consumption, and they try to smooth consumption as well as the rate of change of consumption.

The measure of the external component of the habit follows Abel (1990). It is  $[W_i C_{t-1}]$ , where  $W_i$  is the row vector  $[w_{i1}, \dots, w_{iN}]$ , and  $C_{t-1}$  is the column vector  $[C_{1,t-1}, \dots, C_{N,t-1}]'$ . The external habit can be interpreted as a standard of living that consumers try to achieve. This standard of living,  $W_i C_{t-1}$ , depends on the spatial matrix  $W$ . To define  $W$ , I follow the spatial econometrics literature, and I do not estimate it. Estimating  $W$  using the observed characteristics of the U.S. states is not appropriate because the estimated  $W$  would be endogenous. The exogeneity of  $W$  is essential in avoiding the identification problems elaborated by Manski (1995).

I consider five spatial matrices based on geographical proximity and degree of urbanization. Like Case (1991), Topa (2001), and Ravina (2005), I examine whether the external habit is determined by geographical proximity by considering two specifications for  $W$ . The first measure of external-habit formation assumes that state consumer  $i$  is influenced only by the average consumption of neighboring states. The spatial matrix related to this habit is denoted by  $W_n$ . The second measure is based on gravity models, where the weights in  $W$  are inversely related to the geographic distance between the U.S. states (Hernández-Murillo 2003). In particular, the  $w_{ij}$  weights in the distance-weighted  $W$  (denoted by  $W_d$ ) are

$$w_{ij} = \frac{d_{ij}^{-2}}{\sum_j d_{ij}^{-2}},$$

where  $d_{ij}$  is the distance in miles between the population centroids of state  $i$  and state  $j$ . The population centroids are from the 1990 census. See Hernández-Murillo (2003) for details on the census data.

Whether a state's consumption affects the external-habit level of other states might depend on attributes other than distance. Given the state-level data reported by the Census Bureau, such an attribute can be the degree of urbanization of a state. Recognizing that consumption trends might originate from urban centers, the weights of the third spatial matrix (denoted by  $W_{d,u}$ ) are given by

$$w_{ij} = \frac{d_{ij}^{-2} U_j}{\sum_j d_{ij}^{-2} U_j},$$

where  $U_j$  is the percentage of state  $j$ 's population living in urban areas as reported in the 1990 census. To further investigate the effect of urban centers, the fourth habit measure recognizes that more urban states can be the leaders of consumption trends, and more rural states can be the followers to consumption trends. The fourth measure therefore deletes from the third measure any states with  $U$  less than or equal to 60%. Therefore, the weights of the fourth spatial matrix (denoted by  $W_{d,60}$ ) are given by

$$w_{ij} = \frac{d_{ij}^{-2} U_{j,60}}{\sum_j d_{ij}^{-2} U_{j,60}},$$

where  $U_{60,j} = U_j$  for  $U_j > 60\%$ , and  $U_{60,j} = 0$  for  $U_j \leq 60\%$ . The states that are omitted from the habit measures (i.e. the followers) are Arizona, Georgia, Iowa, Kentucky, Michigan, Mississippi, Nebraska, New Mexico, New York, Pennsylvania, Rhode Island, Texas, and Washington. The hypothesis that the consumption habit depends on the degree of urbanization has not been tested in the existing literature.

Finally, the fifth measure is related to U.S. average consumption growth, which is a popular measure in many asset pricing models (Abel 1990; Campbell and Cochrane 1999; Wachter 2006). In particular, the weights in the spatial matrix of the fifth measure (denoted by  $W_{usa}$ ) are given by  $1/(N - 1)$ . Similar to the previous habit measure, the  $w_{ii}$  weight is set to zero, which implies

that the habit for state  $i$  is based on the consumption of all the states *but* state  $i$ . Since the fifth habit measure is close to U.S. average consumption, I call it U.S. habit. Note that across the five spatial matrices I consider, only the  $W_{usa}$  is symmetric.

## 4.2 Econometric Model

The econometric model is derived from the Euler equations of the state consumers in five steps. First, I obtain the Euler equation for the time-varying risk-free real interest rate,  $R_{it}$ , for the representative consumer of each state:

$$\mathbf{E}_{t-1} [ MU_{i,t-1} - \pi_1 e^{-\beta_i} MU_{it} ] = \mathbf{E}_{t-1} e^{-\beta_i} R_{it} \times [ MU_{it} - \pi_1 e^{-\beta_i} MU_{i,t+1} ], \quad (7)$$

where  $MU_{is}$  is the marginal utility of consumption ( $= (C_{is} - \rho_1 W_i C_{s-1} - \pi_1 C_{i,s-1})^{-\gamma}$ ,  $s = t - 1, t, t + 1$ ). The parameter  $\beta$  is the discount rate, and  $\mathbf{E}_t(\cdot)$  is the conditional expectation at time  $t$ . Following Shapiro (1984), the  $R_{it}$  is the annualized real return of the one-month Treasury bill. Even if the nominal rate is the same across all states, the real rate varies across states because of regional inflation differences.

Second, I follow Deaton (1992) and express the Euler equation (7) as a second-order difference equation in  $MU$ . The solution to this difference equation satisfies

$$MU_{i,t-1} = \mathbf{E}_{t-1} [ e^{-\beta_i} R_{it} MU_{it} ]. \quad (8)$$

Equation (8) holds exactly if the interest rate is constant across time (Hayashi 1985; Ravina 2005). Equation (8) is then a good approximation for assets, like the one-month Treasury bill, which have returns with low time variation (the real return on the one-month Treasury bill has a standard deviation of only 0.027).

Third, I recognize that under rational expectations, the error in forecasting the conditional expectation in the right side of (8) must be uncorrelated with all information available at time



$t - 1$ . In other words,

$$MU_{i,t-1} = e^{-\beta_i} R_{it} MU_{it} + \varepsilon_{it} . \quad (9)$$

As in Shapiro (1984), the random variable  $\varepsilon_{it}$  is a forecasting error and its orthogonal to any information available at  $t - s$ ,  $s > 1$ , i.e. its conditional expectation at time  $(t - s)$  is zero ( $\mathbf{E}_{t-s}\varepsilon_{it} = 0$ ).

Fourth, in line with the empirical consumption and asset pricing literature, I log-linearize equation (9) around the steady state (Breedon, Gibbons, and Litzenberger (1989), Dynan (2000), Fuhrer (2000), Vissing-Jorgensen (2002), Jacobs and Wang (2004), Jagannathan and Wang (2005), Ravina (2005)). The linearization gives rise to an econometric regression, which is a dynamic model of the growth rate of state consumption with fixed and spatial effects:

$$\Delta c_{it} = -\beta_i + \frac{1 - \rho_1 - \pi_1}{\gamma} r_{it} + \rho_1 W_i [\Delta c_{t-1}] + \pi_1 \Delta c_{i,t-1} + \varepsilon_{it} + HT_{it}, \quad (10)$$

where  $HT_{it}$  includes the higher-order terms ignored by the linearization. Also, the lowercase letters denote natural logarithms, and  $\Delta$  denotes first differences, that is,  $\Delta c_t = \ln(C_t) - \ln(C_{t-1})$ . See the Appendix for the details of the log-linearization.

Fifth, I stack the linearized Euler equations (10) of all the state consumers at time  $t$ , and I derive the following pooled dynamic panel model with fixed and spatial effects, as in sections 2 and 3:

$$\Delta c_t = c + \rho_1 [W \Delta c_{t-1}] + \pi_1 [\Delta c_{t-1}] + \lambda_1 r_t + \eta_t. \quad (11)$$

Regression (11) contains the measure of the catching-up component of the habit ( $W \Delta c_{t-1}$ ) and the measure of internal component of the habit ( $\Delta c_{t-1}$ ). A regression parameter  $\rho_1$  of zero yields a pure internal-habit model. A regression parameter  $\pi_1$  of zero yields a pure external-habit model. The parameter  $\lambda_1$  provides an estimate for the elasticity of intertemporal substitution. Given equation (10),  $\lambda_1$  is  $(1 - \rho_1 - \pi_1) / \gamma$ . Moreover, in conjunction with the estimates of  $\rho_1$

and  $\pi_1$ , it can provide an estimate for the curvature parameter  $\gamma$ . The statistical model also includes fixed effects represented by the vector of constants,  $c$ , which originate from the state-specific rates of time preference  $\beta$ . The error term  $\eta$  includes the prediction errors  $\varepsilon$  and the higher-order terms  $HT$  omitted by the log-approximation.

In regression (11), the only control variable is the interest rate  $r_t$ . However as part of the robustness tests, I also estimate regressions with additional explanatory variables like income growth. Carroll (2000) argues that such explanatory variables are endogenous because the error term in a log-linear Euler equation includes higher-order terms, which themselves are contemporaneously correlated with variables like income growth. Therefore, he challenges the use of traditional instrumental variable approaches to estimate (11) since most instruments are also correlated with the error term in a log-linear Euler equation. His criticism does not apply to my estimation technique, which assumes that the  $Z$  instruments are only weakly exogenous (that is,  $Z_t$  is correlated with  $\eta_t$ , but  $Z_{t-1}$  is not).

### 4.3 Empirical Results

This section presents the main empirical findings of the paper. After describing the data, I present the estimates of the habit regression using the biased LSDV. I then estimate the habit regression with the bias-corrected estimator under the five specifications for the spatial matrix  $W$ . I also estimate the habit regression with the instrumental variables approach of Anderson and Hsiao (1982).

*4.3.1 Data Description.* Regression (11) is estimated with annual data for the 1966-98 period for the 48 continental states. My proxy for state consumption is state sales at all retail establishments. Asdrubali, Sorensen, and Yosha (1996), Ostergaard, Sorensen, and Yosha (2002), and Ravina (2005) also use retail sales as their proxy for consumption. The retail sales data are constructed by the Census Bureau from sales tax data, and they are reported in the Statistical Abstract of the United States. Therefore, they suffer from less measurement error than the self-reported individual consumption data from the Consumer Expenditure Survey and the Panel Survey of Income Dynamics. Moreover, my sample ends in 1998 because the Census Bureau

terminated the state-level retail sales program with the 1998 reporting year.

The state retail sales are expressed in real per capita terms. Inflation and population changes are accounted for using the regional price indexes (base year 1992) from the Bureau of Labor Statistics (BLS), and the state population indexes from the Current Population Survey.

The empirical analysis also includes the state income data from the Bureau of Economic Analysis (BEA), which are expressed in real per capita terms. The annual nominal interest rate is compounded from the return of the one-month Treasury bill as reported by the Center for Research in Security Prices. The interest rate is adjusted for inflation using the BLS regional price indexes.

I estimate regression (11) under the five specifications of the spatial matrix  $W$ , discussed in Section 4.1., and I report the empirical results in regressions 1 to 7 in Table 1. In regressions 1 and 2, the external-habit measure is based on the consumption of neighboring states ( $W = W_n$ ). Regression 3 uses the distance-weighted  $W_d$ . Regression 4 uses the distance-weighted and urban-weighted  $W_{d,u}$ . Regressions 5 and 7 use the  $W_{d,60}$  spatial matrix. In regression 6, the habit is given by the average consumption of all states but the state in question ( $W = W_{usa}$ ). Moreover, regression 1 is estimated with the bias LSDV and regression 7 is estimated using the Anderson and Hsiao (1982) estimator in expression (5). Regressions 2 to 6 are estimated with the bias correction estimator  $\hat{\varphi}_c$  in expression (4). To implement  $\hat{\varphi}_c$ , I estimate the bias correction terms using parameter estimates provided by the Anderson and Hsiao (1982) estimator. Also, the instrument  $Z$  in  $\hat{\varphi}_c$  for the interest rate  $r_t$ , the only endogenous control variable in regressions 1 to 7, is its time-lagged value  $r_{t-1}$ . I instrument  $r_t$  with  $r_{t-1}$  because in unreported results (available upon request) I find that it has the highest correlation with  $r_t$  across all the variables in my data set.

*4.3.2 Biased LSDV.* To evaluate the empirical importance of the bias correction, regression 1 is estimated by the biased LSDV from expression (2). In this case, the estimates for  $\rho_1$ ,  $\pi_1$ , and  $\lambda$  are  $-0.02$ ,  $0.29$ , and  $0.28$  respectively. Regression 2 estimates the same model using the bias-corrected estimator, and the estimate for  $\rho_1$ ,  $\pi_1$  and  $\lambda$  become  $0.07$ ,  $0.26$ , and  $0.63$  respectively.

The differences in the estimated parameters translate into important differences in terms of the economic implications of the model. In particular, in regression 1, the curvature parameter  $\gamma$  is insignificant (estimate = 2.60, t-statistic = 1.28), while in regression 2 it is precisely estimated (estimate = 1.05, t-statistic = 8.27). The t-statistic for  $\gamma$  is calculated using the delta method.

*4.3.3 Habit Formation.* The estimation results in regressions 2 to 6 in table 1 reveal that the parameter of the external-habit measure is statistically significant. In particular, when the habit is based on the consumption of neighboring states (i.e.  $W = W_n$ ), the estimate of  $\rho_1$  is 0.26 (t-statistic = 2.50). This estimate increases to 0.33 (t-statistic = 2.61) for the distance-weighted  $W$  and to 0.34 (t-statistic = 2.73) when the degree of urbanization is incorporated in the distance-weighted spatial matrix ( $W = W_{d,u}$ ). The estimated  $\rho_1$  slightly increases when states with a low degree of urbanization are deleted from the distance-weighted and urban-weighted spatial matrix (estimate = 0.37, t-statistic = 3.14). However, the estimate of  $\rho_1$  decreases to 0.34 (t-statistic = 2.12) when the habit is measured by the average consumption of all states but the state in question ( $W = W_{usa}$ ).

The previous results show that the closer two states are, the more they influence one another. Moreover, they suggest that the more urban a state is, the more it seems to influence the other states. The finding that geographical proximity affects the external consumption habit is consistent with existing work (Case 1991; Ravina 2005). The suggestive evidence that urban centers might drive habits is a new result.

Moving to the internal habit, the estimation shows that in the presence of external-habit formation, the internal habit is statistically insignificant—across regressions 2 to 6 the parameter estimate on the time-lagged dependent variable does not exceed 0.07, and its t-statistic does not exceed 1.60. This finding is consistent with existing empirical studies that estimate pure internal-habit models with annual data. For example, Ferson and Constantinides (1991) use national U.S. data and find only weak evidence of internal-habit formation at the annual frequency. Dynan (2000) finds no evidence for internal-habit formation in individual annual consumption data.

Studies that estimate habit models with *quarterly* national data, however, report supporting

evidence for internal-habit models. For example, see Ferson and Constantinides (1991), Ferson and Harvey (1992), Heaton (1995), Grishchenko (2004), and Chen and Ludvigson (2007). Combining the findings with annual and quarterly data, one concludes that the impact of the internal habit must be short-lived because it manifests itself at the quarterly frequency and it disappears at the annual frequency.

One reason for failing to accept the internal-habit formation assumption may be the fact that my measure of internal-habit formation does not include many consumption lags. Therefore, I include an additional time lag of own consumption growth,  $\Delta c_{t-2}$ , to regression 5. In untabulated results, available on request, the parameter estimates for the internal-habit terms  $\Delta c_{t-1}$  and  $\Delta c_{t-2}$  are again statistically insignificant. In contrast, the parameter estimate on the external component of the habit is precisely estimated. See the Appendix on how to estimate a model with higher order lag-dependent variables.

*4.3.4 Elasticity of Intertemporal Substitution.* The coefficient on the log real interest rate provides an estimate for the elasticity of intertemporal substitution (EIS). The estimate of the EIS is significantly different from zero at the 5% confidence level in regressions 2 to 6. Its value is positive and around 0.61. This estimate is higher than the estimates found using U.S. national data (Hall 1988) and is in line with estimates from studies using individual data (Zeldes 1989; Attanasio and Weber 1993, 1995; Vissing-Jorgensen 2002). Beaudry and Wincoop (1996) also report that the Treasury bill return has a significant effect on state consumption growth.

The estimates on the habit measures and the EIS provide an estimate for the curvature parameter  $\gamma$  because the EIS equals  $(1 - \pi_1 - \rho_1)/\gamma$ . Using the last expression, the implied estimate of  $\gamma$  is around 1.01 across regressions 2 to 6. The estimated  $\gamma$  is also statistically significant at the 5% level (the standard error for  $\gamma$  is calculated using the delta method). The estimate of  $\gamma$  implies that the coefficient of relative risk aversion, which is given by  $\gamma/S_{it}$ ,  $S_{it} = 1 - (\rho_1 W_i C_{t-1} + \pi_1 C_{i,t-1})/C_{it}$ , is on average 1.60 across regressions 2 to 6. This value is reasonable and within the range used in the literature. For example, see Grossman and Shiller (1981), and Campbell and Cochrane (1999).

**4.3.5 Instrumental Variables.** An alternative to the bias-correction estimation is the instrumental variables approach, which produces similar results to the bias-corrected estimator. Following Anderson and Hsiao (1982), I factor out the fixed effects by taking first differences of the linearized Euler equation in (11):  $\Delta^2 c_t = \rho_1 [W \Delta^2 c_{t-1}] + \pi_1 [\Delta^2 c_{t-1}] + \lambda_1 \Delta r_t + \Delta \eta_t$ , where  $\Delta^2$  denotes second order differences ( $\Delta^2 x_t = x_t - 2x_{t-1} + x_{t-2}$ ). Also, the instruments for  $W \Delta^2 c_{t-1}$ ,  $\Delta^2 c_{t-1}$ , and  $\Delta r_t$ , are  $W \Delta c_{t-2}$ ,  $\Delta c_{t-2}$ , and  $r_{t-2}$  respectively. The Anderson and Hsiao (1982) estimates are reported in regression 7. As in regression 5,  $\rho_1$  is significant (estimate = 0.60, t-statistic = 2.64) and  $\pi_1$  is insignificant (estimate = -0.03, t-statistic = -0.75). Also, the estimate of the EIS (= 0.91, t-statistic = 1.50) implies a statistically significant estimate for  $\gamma$  (= 0.48, t-statistic = 2.00).

Because both the Anderson and Hsiao (1982) and bias-corrected estimators are consistent, the similarity of their qualitative results is not surprising. However, the bias-corrected estimator provides more precise estimates for the  $\rho_1$ ,  $\gamma$ , and EIS. This finding is confirmed in the simulation results in section 5, in which the standard deviation of the Anderson and Hsiao (1982) estimates are always higher than the standard deviation of the bias-corrected estimates.

#### 4.4 Robustness Tests

Overall, the evidence for internal-habit formation is weak. However, the evidence for external-habit formation based on geographical proximity and degree of urbanization is more substantial. I test the robustness of these conclusions by focusing on the distance-weighted and urban-weighted  $W$  that excludes states with a low degree of urbanization (table 1, regressions 8-14).

These regressions are estimated with the bias corrected LSDV. To implement  $\hat{\varphi}_c$ , I estimate the bias correction terms using parameter estimates provided by the Anderson and Hsiao (1982) estimator. All regressions include the interest rates as a control variable and the instrument  $Z$  for the interest rate  $r_t$  is its time-lagged value  $r_{t-1}$ .

Besides  $r_t$ , the regressions 8 to 12 include an additional control variable. In particular, regression 8 includes the growth rates of U.S. services. Its instrument in  $\hat{\varphi}_c$  is the average growth rate of the past state income  $W_{d,60} \Delta inc_{t-1}$  ( $\Delta inc_{t-1}$  denotes the growth rate of lagged state

income). Regression 9 includes the growth rate of state income,  $\Delta inc_t$ , and its instrument in  $\hat{\varphi}_c$  is  $W_{d,60}\Delta inc_{t-1}$ . Regression 10 adds the growth rate of time-lagged state income,  $\Delta inc_{t-1}$ , which is instrumented in  $\hat{\varphi}_c$  by  $\Delta inc_{t-1}$ . Regression 11 includes the squared consumption growth rate and its instrument in  $\hat{\varphi}_c$  is the average squared time-lagged consumption growth rate across all states but the state in question. Regression 12 includes the average income growth rate of the states that determine the habit,  $W_{d,60}\Delta inc_t$ . Its instrument in  $\hat{\varphi}_c$  is  $W_{d,60}\Delta inc_{t-1}$ . I use the aforementioned instruments because in unreported results (available upon request), I find that they are highly correlated with the variables they are instrumenting.

*4.4.1 Services.* State consumption growth is measured with error because it is equal to the growth rate of state retail sales, which excludes the growth rate of services. Thus, the growth rate of services at the state level is an omitted variable. To account for this omission, regression 8 includes the growth rate of U.S. services. Because the growth rate of U.S. services should be correlated to the consumption of services at the state level, it can alleviate potential omitted variable biases. Conditional on the growth rate of U.S. services, I find that the external-habit term remains significant (estimate = 0.40, t-statistic 2.58), the internal-habit term is insignificant (estimate = 0.02, t-statistic = 0.57), and the EIS is significant (estimate = 0.92, t-statistic = 3.43). The issue of measurement error is further investigated in the simulation analysis in section 5.2.

*4.4.2 Income Expectations.* The literature finds that consumption growth is excessively sensitive to income growth, that is, current consumption growth is related to current income growth (Flavin 1981, 1993; Campbell and Mankiw 1989). One reason for this phenomenon is that current income growth can help predict future income growth, which influences current consumption growth. It is therefore possible that the external-habit measure is significant because it captures some common dimension of economic activity that helps agents forecast their future income. To control for the excess sensitivity of consumption growth to income growth, regression 9 includes state income growth as an additional regressor. This estimation shows that the external component of the habit remains significant in the presence of state income growth ( $\rho_1$  estimate

$= 0.42$ ,  $\rho_1$  t-statistic  $= 2.11$ ).

*4.4.3 Liquidity Constraints.* The economic model in section 4.1 ignores liquidity constraints. If liquidity constraints bind simultaneously across all U.S. states, because of a negative national income shock, consumption growth rates across all states will decrease. If the negative income shock is severe (persistent), state consumption levels might remain low for more than one period. In this case, the low consumption growth rate of a state will be correlated with the low past consumption growth rates of other state in the absence of external-habit formation. I find that this is not the case in my estimation. In line with Zeldes (1989) and Deaton (1991, 1992), I account for liquidity constraints by adding the growth rate of time-lagged state income in regression 10. In this regression,  $\rho_1$  is 0.40 and it remains statistically significant (t-statistic  $= 3.02$ ).

*4.4.4 Precautionary Savings.* Carroll (1997) argues that individuals save due to precautionary motives to finance consumption in unforeseen bad states of the economy. As Dynan (1993) explains, in the presence of precautionary motives, “higher expected consumption growth (which reflects higher saving) is associated with higher expected squared consumption growth (which reflects greater uncertainty)” (page 1106).

Following Dynan (1993), I measure consumption uncertainty by the squared consumption growth rate, which I add to the estimation in regression 11. Like Dynan (1993) and Ravina (2005), I find that the squared consumption growth rate is insignificant. In addition, as in regression 5, the coefficient on the external habit remains significant (estimate  $= 0.31$ , t-statistic  $= 2.17$ ). These findings suggest that the external-habit measure is not significant because it captures some dimension of overall consumption uncertainty.

*4.4.5 Long-Run Convergence.* Barro and Sala-i-Martin (1992) find that the income of U.S. states is converging, that is, the income of states that had low past income tends to grow faster than the income of states with high past income. Even though income convergence creates convergence in U.S. state consumption, income convergence does not spuriously inflate the importance of the external-habit measures.

To account for the effects of income convergence, one can, like Barro and Sala-i-Martin (1992),



include the level of state income at the beginning of the sample period as an additional control variable. Because the initial value of income is not time-varying, it is eliminated from the model when I de-mean the data prior to the estimation. Consequently, the significance of the external-habit measure is not a by-product of long-run income convergence.

*4.4.6 Common Shocks.* Next, I consider the effect of unobserved common shocks on the significance of the external-habit measures. When common shocks are independent across time, they create cross-sectional correlations in the error term. When they are autocorrelated, they also create autocorrelation in the error term. In both cases, the error term is not IID and condition 1(1) is violated.

Since the corrected estimator cannot account for the presence of common shocks directly, I modify regression (11) in one of two ways to factor out the impact of common shocks. First, I use control variables that should be correlated with the common shocks. Second, I estimate regression (11) with modified (relative) state consumption data that are less influenced by common shocks compared to actual state consumption data.

I use income growth rates to purge the effects of common shocks from the error terms because income growth determines consumption growth, and thus unobserved income shocks are the source of unobserved consumption shocks. In particular, regression 12 adds  $W_{d,60}\Delta inc_t$  to the estimation. The variable  $W_{d,60}\Delta inc_t$  is the average income growth rate of the states that are *included* in the definition of the external-habit measure. In this regression, the coefficient estimate on the external-habit measure is 0.39, and its statistical significance (t-statistic = 2.06) is not affected by the presence of  $W_{d,60}\Delta inc_t$ . Also, recall that in regression 9 the coefficient estimate on the external-habit term is significant even in the presence of the state's *own* income growth rate, which should also be affected by unobserved macroeconomic shocks.

A different way to factor out common shocks is to estimate the baseline regression in (11) by replacing consumption growth with relative consumption growth  $\Delta c_R$  (regression 13). Similar to Ostergaard, Sorensen, and Yosha (2002), relative consumption growth,  $\Delta c_R$ , is the difference between state consumption growth and state income growth. Because income growth includes

any common unobserved shocks that matter for consumption growth, the variation in relative consumption growth is primarily state-specific. In this case, the estimation shows that the relative habit growth  $W_{d,60}\Delta c_{R,t-1}$  is statistically significant in explaining relative consumption growth  $\Delta c_R$  ( $\rho_1$  estimate = 0.34,  $\rho_1$  t-statistic = 3.19). In an additional robustness test, I estimate a regression in which relative consumption is the difference between state consumption and U.S. income. Untabulated results show that  $\rho_1$  is again precisely estimated ( $\rho_1$  estimate = 0.36,  $\rho_1$  t-statistic = 3.61).

*4.4.7 Cross-Border Shopping.* In a final robustness test, I find that potential cross-border shopping has no impact on the significance of the external component of the habit. Cross-border shopping can cause spending in two neighboring states to be positively correlated, independent of any external-habit effects. For example, a positive income shock in Connecticut increases the purchasing power, and subsequent consumption, of Connecticut residents. It is reasonable to assume that some of this excess purchasing power will be spent in neighboring states such as New York. To factor out the effect of cross-border shopping, I exclude the neighboring states from the distance-based urban-based external-habit measure that omits states with low degree of urbanization. In regression 14, the parameter estimate on the external-habit term is again precisely estimated (estimate = 0.36, t-statistic = 2.86).

To summarize, the estimation results find that the external-habit measures based on geographical proximity and degree of urbanization are the most statistically significant. The estimates of these external-habit measures remain significant in the presence of various control variables like state income growth rates. Also, external-habit formation does not reflect alternative theories that include liquidity constraints, precautionary savings, and the long-run convergence of state income.

## 5. FINITE-SAMPLE PERFORMANCE

Having estimated the regression with habit formation, I conduct a simulation exercise based on the empirical results in section 4. The simulations analyze the finite-sample properties of the

bias-corrected estimator  $\widehat{\varphi}_c$  with and without measurement error in the dependent variable. The case of measurement error is relevant because in many empirical applications the true value of the dependent variable is unobserved. I also use the simulations to compare the bias-corrected estimator  $\widehat{\varphi}_c$  to the pure instrumental variable estimator  $\widehat{\varphi}_{AH}$  of Anderson and Hsiao (1982), which is the main alternative to  $\widehat{\varphi}_c$ .

### 5.1 Simulation Set-Up

In the case of no measurement error, the data-generating mechanisms (DGM) of the dependent variable  $Y_t$  and the endogenous control variable  $X_t$  are

$$Y_t = \pi_1 Y_{t-1} + \rho_1 (WY_{t-1}) + \lambda X_t + c + \eta_t, \text{ and} \quad (12a)$$

$$X_t = a_1 X_{t-1} + a_0 + \eta_{x,t}, \quad (12b)$$

where  $\eta_t = \zeta (\varepsilon_{com,t} + \varepsilon_t)$ , and  $\eta_{x,t} = \varepsilon_{x,t} + a_{\eta x} \varepsilon_{com,t}$ . The error terms  $\varepsilon_{x,t}$ ,  $\varepsilon_{com,t}$ , and  $\varepsilon_t$  are all IID normally distributed with a mean of zero and a variance of 1. Also,  $\pi_1 = 0.03$ ,  $\rho_1 = 0.37$ ,  $\lambda = 0.59$ ,  $c \sim \text{NIID}(0.01, 1)$ ,  $a_1 = 0.78$ ,  $a_0 \sim \text{NIID}(0.014, 1)$ ,  $a_{\eta x} = 0.33$ , and  $\zeta = 2.01$ .

In the case of  $Y_t$ , the values of  $\pi_1$ ,  $\rho_1$  and  $\lambda$  are from table 1, regression 5. The vector of fixed effects,  $c$ , is normally distributed and its mean is the cross-sectional average of the mean consumption growth rate across the U.S. states. The matrix  $W$  is the distance-weighted and urban-weighted spatial matrix that ignores states with a low degree of urbanization. i.e.  $W = W_{d,60}$ .

As in regression 5 in table 1, the simulation has only one control variable  $X_t$ , which is defined to reflect the properties of the log real interest rate. The DGM of  $X_t$  follows the pooled dynamic AR(1) process in (12b). The vector  $a_0$  contains the fixed effects of  $X$ , which are drawn from a normal distribution. Its mean is the cross-sectional average of the mean log real interest rate across the U.S. states. The value for  $a_1$  is the estimate from an AR(1) pooled dynamic panel model with fixed effects on the log real interest rate. The AR(1) is estimated using the procedure in Hahn and Kuersteiner (2002).

The error terms  $\eta_t (= \zeta (\varepsilon_{com,t} + \varepsilon_t))$  and  $\eta_{x,t} (= \varepsilon_{x,t} + a_{\eta x} \varepsilon_{com,t})$  are correlated because they

both include the common shock  $\varepsilon_{com,t}$ . Also, the parameter  $\zeta$  in  $\eta_t$ , and the parameter  $a_{\eta x}$  in  $\eta_{x,t}$  are chosen to match two facts: (a) the ratio of the variances (*RATIO*) of the log real interest rate and estimated residual  $\hat{\eta}$  from regression 5 in table 1, and (b) the correlation between  $X_t$  and  $\eta_t$  (*COR<sub>x $\eta$</sub>* ). Then, using the expressions for  $Y_t$ ,  $\eta_t$ , and  $\eta_{x,t}$ ,

$$a_{\eta x} = \left[ 2 \frac{1 - a_1^2}{COR_{x\eta}^2} - 1 \right]^{-1/2} = 0.33, \text{ and } \zeta = \left[ \frac{1 + a_{\eta x}^2}{2(1 - a_1^2) RATIO} \right]^{1/2} = 2.01,$$

because in regression 5, the levels of *RATIO* and *COR<sub>x $\eta$</sub>*  are 0.35 and 0.28 respectively.

Using the above set-up, four cases are simulated under four values of  $T$ : 20, 30, 40, and 60. In each case, 20,000 draws are executed. Each draw generates  $T + 100$  observations for each cross-sectional unit. The first 100 observations are deleted to mitigate the effects of the initial conditions. For every draw, the sample of  $T$  observations is used to estimate  $\pi_1$ ,  $\rho_1$ , and  $\lambda$ .

Three estimators are considered. The first estimator is the biased LSDV from expression (2). The second is the bias-corrected  $\hat{\varphi}_c$ , which is implemented in the way described in section 3.3. The third is the Anderson and Hsiao (1982) instrumental variables estimator  $\hat{\varphi}_{AH}$  from expressions (5). In  $\hat{\varphi}_{AH}$ , the instruments for  $\Delta Y_{t-1}$ ,  $W\Delta Y_{t-1}$ , and  $\Delta X_t$ , are  $Y_{t-2}$ ,  $WY_{t-2}$ , and  $X_{t-2}$  respectively. For each of the three estimators, the percentage mean bias, standard error, and root mean squared error (RMSE) of  $\pi_1$ ,  $\rho_1$ , and  $\lambda$  are calculated with their estimates from the 20,000 draws (table 2, with no measurement error; table 3, with measurement error). In what follows, I denote the Anderson and Hsiao (1982) estimator by AH.

## 5.2 Results with No Measurement Error

When the dependent variable is accurately measured, the LSDV is severely biased, but it has a low standard error and RMSE. The bias-corrected  $\hat{\varphi}_c$  has a low finite-sample bias, together with a low standard error and RMSE. The AH estimator has the smallest bias. Because it has the highest standard error however, its RMSE is higher than the RMSE of  $\hat{\varphi}_c$ .

The previous findings are drawn from the simulation results reported in table 2. Consider the case with  $T = 30$ . For  $\rho_1$ , the percentage bias is the highest for LSDV ( $-30.4$ ), and it is the smallest for the AH ( $-21.9$ ). It is also only  $-27.1$  for the bias-corrected estimator. However,

in terms of the variability of  $\rho_1$ , the LSDV has the smallest standard error (6.0), and the AH has the highest standard error (10.1). The standard deviation for the bias-corrected estimator is only 6.2, and it is very close to that of the LSDV. Finally, for  $\rho_1$ , the  $\widehat{\varphi}_c$  has the smallest RMSE (11.8) and the AH has the highest RMSE (12.9). Similar results hold for  $\pi_1$  and  $\lambda$ .

The findings of the simulation exercise reveal that the bias-corrected estimator retains the low variance of the uncorrected LSDV, while it has a significantly lower bias. The low variance of the LSDV is the motivation for Kiviet (1995) to pursue the bias correction approach in the context of dynamic panel models with no spatially lagged dependent variables and no endogenous control variables. In Kiviet (1995), as in the present study, the bias-corrected estimator has comparable finite-sample bias but has smaller standard deviation than pure instrumental variable estimators.

### 5.3 Results with Measurement Error

A set of simulation results quantify the effect of measurement error on the finite-sample performance of the bias-corrected estimator and the AH estimator. In these simulations, the observed value of  $Y$ ,  $Y_{obs}$ , is

$$Y_{obs} = Y^* + v, \quad v \sim \text{NIID}(0, \sigma_v^2),$$

where  $Y^*$  is the true value of  $Y$  and  $v$  is the measurement error. The variance of  $v$  is  $\sigma_v^2$ , and its level is chosen so that the signal-to-noise ratio is 1 for all cross-sectional units. Also, the measurement error is orthogonal to error terms  $\eta$  (from equation (12a)) and  $\eta_x$  (from equation (12b)). To focus on the effect of measurement error in  $Y$ , I assume that  $X$  is free of measurement error ( $X_{obs} = X^*$ ). The values for  $Y^*$  and  $X^*$  are generated as in the case of no measurement error presented in section 5.1.

The simulation results for the  $\widehat{\varphi}_c$  and the AH estimators are of two types (table 3). In the type 1 simulations, the instruments for the AH do not suffer from measurement error, that is the instruments for  $\Delta Y_{obs,t-1}$ ,  $W\Delta Y_{obs,t-1}$ , and  $\Delta X_{obs,t}$ , are  $Y_{t-2}^*$ ,  $WY_{t-2}^*$ , and  $X_{t-2}^*$  respectively. In practice of course, it might be almost impossible to find such instruments. Nevertheless, the type 1 simulations can illustrate whether the corrected LSDV can perform better than the AH,

which uses the best possible instruments, i.e. instrument with no measurement error. In the type 2 simulations, the instruments for the AH are subject to measurement error, that is the instruments for  $\Delta Y_{obs,t-1}$ ,  $W\Delta Y_{obs,t-1}$ , and  $\Delta X_{obs,t}$ , are  $Y_{obs,t-2}$ ,  $WY_{obs,t-2}$ , and  $X_{t-2}^*$  respectively. Because the bias vector in  $\hat{\varphi}_c$  is computed using the AH, the value of  $\hat{\varphi}_c$  is different between the type 1 and type 2 simulations. The simulation results for  $\lambda$  are similar to those reported in table 2, and are omitted from table 3.

For  $\pi_1$ , in the type 1 simulations, the  $\hat{\varphi}_c$  is more biased than AH. The result is not surprising because in the presence of measurement error the  $\hat{\varphi}_c$  is inconsistent, while the AH is consistent. However, the bias-corrected estimator has a smaller standard deviation and RMSE than the AH that uses instruments, which are free of measurement error. Moving to the type 2 simulations, the AH estimate of  $\pi_1$  has a high percentage bias because the instruments for AH are subject to measurement error. Also, the standard error and RMSE of the AH are higher than those of the bias-corrected estimator.

For  $\rho_1$ , in the type 1 simulations, the percentage finite-sample bias for the bias-corrected estimator is higher than the AH, and its standard deviation is lower than the AH. However, its RMSE is slightly higher than the AH. Moving to the type 2 simulations, the bias-corrected estimator now has a smaller percentage bias, standard deviation, and RMSE than the AH.

Overall, the results show that in the presence of measurement error, the bias-corrected estimator maintains fairly good finite-sample properties. Also, compared to the AH, it always has a smaller standard error.

## 6. CONCLUSION

A leading explanation of aggregate stock market behavior states that asset returns are determined as if consumers' welfare depends on the difference between actual consumption and a habit level of consumption. In some models, the habit is determined by the consumer's past consumption (internal-habit formation). In other models, it is determined by the consumption of others (external-habit formation). To test for the relative strength of these two types of habit

formation, this paper proposes an estimator for dynamic panel models with fixed effects, spatial effects, and endogenous regressors.

The estimator is based on the bias-correction method of Hahn and Kuersteiner (2002) for dynamic panel models with only fixed effects. I modify their approach to accommodate spatial effects and endogenous control variables. The proposed bias-corrected estimator is a hybrid of the least-squares dummy variable estimator of Hahn and Kuersteiner (2002) and the instrumental variables estimator of Anderson and Hsiao (1982). Like Hahn and Kuersteiner (2002), I use de-meaned data to eliminate the fixed effects, and like Anderson and Hsiao (1982), I instrument the endogenous control variables. I show that the new estimator is consistent, has no asymptotic bias, and is asymptotically normally distributed.

I apply the estimator to consumption data for the 48 continental U.S. states to test for habit formation at the aggregate level. The estimation considers both internally and externally determined habits. The internal-habit level is measured by own time-lagged consumption growth. The external-habit level is measured by various weighted averages of time-lagged consumption growth rates of states other than the state in question. The estimation finds weak evidence for internal-habit formation, and strong evidence of external-habit formation. In particular, the farther apart two states are, the less they influence one another. Furthermore, states with population that predominantly lives in urban centers seem to affect the consumption of other states the most. The finding that geographical proximity affects the external consumption habit is consistent with existing work (Case 1991; Ravina 2005). The suggestive evidence that consumption trends might originate from urban centers is a new result.

Finally, the bias-corrected estimator has good finite-sample properties. Simulation results show that it has a low finite-sample bias and a low standard error and RMSE. The pure instrumental variable estimator by Anderson and Hsiao (1982) exhibits a smaller bias, but because it has the highest standard error, its RMSE is higher than the RMSE of the bias-corrected approach.

## APPENDIX

### A.1 General Results

The norm used in the appendix is the maximum row sum norm defined next.

**Definition: Norm.** Let  $A$  be an  $N \times N$  matrix and  $a_{ij}$  be the element on its  $i^{th}$  row and  $j^{th}$  column. The maximum row sum norm of  $A$  is  $\|A\| = \max_i \sum_{j=1}^N |a_{ij}|$ ,  $1 \leq i \leq N$ . If  $A$  is an  $N \times 1$  vector, then its norm is  $\|A\| = \max_i |a_i|$ .

The properties of the matrix  $[I - [\rho_1 / (1 - \pi_1)] W]$ , which appears in many of the proofs in Sections A.2 and A.3, are presented in Lemma 1.

**Lemma 1.** Under conditions 1(3) and 2, the matrix  $[I - [\rho_1 / (1 - \pi_1)] W]$  is an M-matrix, it is nonsingular, and the maximum row sum norm of its inverse is bounded by  $[1 - |\rho_1 / (1 - \pi_1)|]^{-1}$ .

**Proof for Lemma 1.** In general, a real square matrix  $D$  of order  $N$  is called an M-matrix if  $D = (rI - A)$ , in which  $I$  is the identity matrix,  $A$  is a nonnegative matrix, and  $r$  is a real number. Also,  $r$  has to be greater than the absolute value of the maximal eigenvalue of the matrix  $A$  denoted by  $e_{|\max|}(A)$ . See Ibragimov (2000), definition 1. By Frobenius' theorem (Minc 1988, chapter 2) we know that the maximal eigenvalue of any non-negative matrix  $A$  is bounded by the maximum row sum of  $A$ . Also note that in my case,  $W$  is by construction a nonnegative matrix, and, by condition 2(3), its maximum row sum is 1.

Define  $D^* = [I - [\rho_1 / (1 - \pi_1)] W]$  as a reformulation of the matrix  $D$ :

$$D^* = \left( \frac{\rho_1}{1 - \pi_1} \right) \left( \frac{1 - \pi_1}{\rho_1} I - W \right) = \frac{1}{r} (rI - W) = \frac{1}{r} D \quad \text{with} \quad \frac{\rho_1}{1 - \pi_1} = \frac{1}{r}.$$

Thus,  $D^*$  is a scaled version of  $D$ , and  $\rho_1 / (1 - \pi_1)$  is the scaling factor. The matrix  $D^*$  is therefore an M-matrix when  $|r|$  is greater than  $e_{|\max|}(W)$ . Because the rows of  $W$  sum to one, its maximum eigenvalue is bounded by 1. Thus, as long as  $|r| > 1$ ,  $D^*$  is an M-matrix. Since  $\rho_1 / (1 - \pi_1) = 1/r$ , the condition  $|r| > 1$  translates to  $|\rho_1| / |1 - \pi_1| < 1$ . The ratio of  $|\rho_1|$  to  $|1 - \pi_1|$  is less than 1 because Condition 1(3),  $0 < |\rho_1| + |\pi_1| < 1$ , implies that

$$|\rho_1| + |\pi_1| < 1 \Rightarrow |\rho_1| < 1 - |\pi_1| \leq |1 - \pi_1| \Rightarrow |\rho_1| < |1 - \pi_1| \Rightarrow |\rho_1 / (1 - \pi_1)| < 1.$$

Therefore,  $D^*$  is an M-matrix.

Next, I find the row sum norm of  $D^*$ . Since  $D^*$  is an M-matrix, the Stolper-Samuelson condition implies that it is nonsingular and all the elements of its inverse are nonnegative (Ibragimov 2000, statement VII). Also, one can bound the norm of the inverse of  $D^*$  (Ibragimov 2000, statement IX):

$$\|(D^*)^{-1}\| = \left\| \sum_{i=0}^{\infty} \left( \frac{\rho_1}{1 - \pi_1} W \right)^i \right\| \leq \sum_{i=0}^{\infty} \left| \frac{\rho_1}{1 - \pi_1} \right|^i \times \|W\|$$

Because the row sum norm of  $W$  is one, the above expression simplifies to:

$$= \sum_{i=0}^{\infty} \left| \frac{\rho_1}{1 - \pi_1} \right|^i = \frac{1}{1 - |\rho_1 / (1 - \pi_1)|}.$$



Finally, I establish two important results for  $\Psi^s = (\pi_1 I + \rho_1 W)^s$ , where  $I$  is the  $N$  by  $N$  identity matrix and  $W$  is the spatial matrix:

**Result 1.** The maximum row sum norm of  $\Psi^s$  is  $k_\Psi$  and it is strictly less than one because

$$k_\Psi = \|\Psi^s\| = \|(\pi_1 I + \rho_1 W)^s\| \leq \|\pi_1 I + \rho_1 W\|^s \leq [|\rho_1| + |\pi_1|]^s < 1. \quad (\text{A.1})$$

The above calculations use the facts that (a) the row sum norms of  $I$  and  $W$  is one, and (b) Condition 1(3) prescribes that  $|\rho_1| + |\pi_1| < 1$ . Moreover, the norm of  $\Psi^s (\Psi^s)'$  is  $k_{\Psi\Psi}$  and it also strictly less than 1:

$$k_{\Psi\Psi} = \|\Psi^s (\Psi^s)'\| = (\|\Psi\| \|\Psi'\|)^s \leq [ (|\rho_1| + |\pi_1|) (|\pi_1| + |\rho_1| \|W'\|) ]^s.$$

Because by Condition 2(4) the maximum column sum of  $W$  is  $k_c$ , and  $k_c$  is less than  $k_0$ ,  $k_0 = \frac{1}{|\rho_1|} \left( \frac{1}{|\rho_1| + |\pi_1|} - |\pi_1| \right)$ , the row sum norm of  $W'$  is also strictly bounded by  $k_0$ ,

$$\begin{aligned} &< \left\{ [|\rho_1| + |\pi_1|] \left[ |\pi_1| + |\rho_1| \frac{1}{|\rho_1|} \left( \frac{1}{|\rho_1| + |\pi_1|} - |\pi_1| \right) \right] \right\}^s \\ &= \left\{ [|\rho_1| + |\pi_1|] \left[ |\pi_1| + \frac{1}{|\rho_1| + |\pi_1|} - |\pi_1| \right] \right\}^s \\ &= \left\{ [|\rho_1| + |\pi_1|] \frac{1}{|\rho_1| + |\pi_1|} \right\}^s = 1 \end{aligned}$$

**Result 2.** The maximum row sum norm of  $\sum_{s=0}^{t-1} \Psi^s$  is bounded,

$$\left\| \sum_{s=0}^{t-1} \Psi^s \right\| = \sum_{s=0}^{t-1} \|\Psi^s\| \leq \sum_{s=0}^{t-1} [|\rho_1| + |\pi_1|]^s \leq \frac{1 - [|\rho_1| + |\pi_1|]^t}{1 - |\rho_1| - |\pi_1|}. \quad (\text{A.2})$$

Moreover, because by Condition 1(3)  $|\rho_1| + |\pi_1| < 1$ , the norm  $\left\| \sum_{s=0}^{t-1} \Psi^s \right\|$  converges to  $1/(1 - |\rho_1| - |\pi_1|)$  as  $t$  goes to infinity. Similarly,  $\left\| \sum_{s=0}^{t-1} \Psi^s (\Psi^s)' \right\|$  is bounded and it converges to  $1/(1 - k_{\Psi\Psi})$  as  $t$  goes to infinity.

## A.2 Approximate Score Function

Like Hahn and Kuersteiner (2002) (HK), I derive the bias-correction terms of the hybrid estimator using the expectation of an approximation to the score function. To obtain the approximate score function, I first rewrite the model in matrix form:

$$Y_t = (\pi_1 I + \rho_1 W) Y_{t-1} + X_t \lambda + c + \eta_t.$$

By recursive substitution, I solve it backwards to obtain its long-run representation:

$$Y_t = (\pi_1 I + \rho_1 W)^t Y_0 + \sum_{s=0}^{t-1} (\pi_1 I + \rho_1 W)^s (c + X_{t-s} \lambda + \eta_{t-s}).$$

In line with HK, I use the long-term representation of  $Y_t$  to define an approximation  $u_t^{**}$  to  $Y_t$ . The

approximation  $u_t^{**}$ , unlike  $Y_t$ , is initialized in the infinite past. In particular, for  $t \geq 1$ :

$$u_t^{**} = \sum_{s=0}^{\infty} \Psi^s (c + X_{t-s}\lambda + \eta_{t-s}) = \Psi^t u_0^{**} + \sum_{s=0}^{t-1} \Psi^s (c + X_{t-s}\lambda + \eta_{t-s}), \quad (\text{A.3})$$

where  $\Psi^s = (\pi_1 I + \rho_1 W)^s$ , and  $u_0^{**}$  is a random variable. Each  $u_{0,i}^{**}$  represents the infinite past history of  $u_{it}^{**}$ . The properties of  $u_0^{**}$  are described in condition 5:

*Condition 5.* The variable  $u_0^{**}$  is an  $N \times 1$  vector given by  $(u_{0,1}^{**}, \dots, u_{0,N}^{**})'$ , which is independent of  $\eta_t$  for all  $t$ . All the moments of  $u_0^{**}$  exist and  $E |u_{0,i}^{**}|^{2+\zeta} < \infty$  for some  $\zeta > 0$  and all  $i$ . The variance of  $u_{0,i}^{**}$  is  $\sigma_{u_0}^2$  for all  $i$ .

Next, I establish the following lemma for  $u_t^{**}$ .

**Lemma 2.** Under conditions 1 to 5, all the moments of the random variable  $u_{it}^{**}$ ,

$$u_{it}^{**} = \sum_{s=0}^{\infty} \Psi_i^s (c + X_{t-s}\lambda + \eta_{t-s}) = \Psi_i^t u_0^{**} + \sum_{s=0}^{t-1} \Psi_i^s (c + X_{t-s}\lambda + \eta_{t-s}),$$

exist. The row vector  $\Psi_i^s$  represents the  $i^{th}$  row of the matrix  $\Psi$  raised to the power of  $s$ , i.e.  $\Psi^s = (\pi_1 I + \rho_1 W)^s$ .

**Proof of Lemma 2.** The random variable  $u_{it}^{**}$  is a linear combination of the  $c$ ,  $X_{t-s}$ , and  $\eta_{t-s}$ . Under conditions 1 and 2, all the moments of  $X_{t-s}$  and  $\eta_{t-s}$  exist. Therefore, as long as the row vector  $\Psi_i^s$  is absolutely summable, the moments of  $u_{it}^{**}$  will be bounded. Under condition 1, one can show that

$$\left\| \sum_{s=0}^{\infty} \Psi_i^s \right\| \leq \sum_{s=0}^{\infty} \|\Psi_i^s\| \leq \sum_{s=0}^{\infty} \|\Psi^s\| = \frac{1}{1 - |\rho_1| - |\pi_1|} < \infty.$$

The last line in the above display follows from result (A.2) when  $t \rightarrow \infty$  because under Condition 1(3),  $|\rho_1| + |\pi_1| < 1$ .

Using  $u_t^{**}$ , I follow HK and approximate the sample score function of the hybrid estimator  $\hat{\varphi}_b$  by replacing the data vector  $\tilde{Z}_{i,t-1}^d$  with a new vector  $u_{it}^{**}$ :

$$S_{NT}^* = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (u_{i,t-1}^{**})' (\eta_{it}^d), \quad (\text{A.4})$$

where  $u_{i,t-1}^{**}$  is the  $1 \times (K+2)$  vector  $(u_{i,t-1}^{**}, W_i u_{i,t-1}^{**}, Z_{i,t-1})$ . Now, we can prove the following two lemmas.

**Lemma 3.** Under conditions 1 to 5, the  $S_{NT}$  score function of  $\hat{\varphi}_b$  can be approximated up to an  $o_p(1)$  with  $S_{NT}^*$ .

**Proof of Lemma 3.** The score function  $S_{NT}$  of  $\hat{\varphi}_b$  can be decomposed in two parts:

$$\frac{1}{\sqrt{NT}} \sum_{i,t} \left( \tilde{Z}_{i,t-1} - \bar{\tilde{Z}}_i \right)' (\eta_{it} - \bar{\eta}_i) = \frac{1}{\sqrt{NT}} \sum_{i,t} \tilde{Z}_{i,t-1} (\eta_{it} - \bar{\eta}_i) - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \bar{\tilde{Z}}_i' \sum_{t=1}^T (\eta_{it} - \bar{\eta}_i),$$

where  $\sum_{i,t}$  denotes a double sum over  $i$  and  $t$ . The variables  $\bar{\tilde{Z}}_i (= T^{-1} \sum_{t=1}^T \tilde{Z}_{it})$  and  $\bar{\eta}_i (= T^{-1} \sum_{t=1}^T \eta_{it})$  denote the average values of  $\tilde{Z}$  and  $\eta$  over  $t$ , respectively. Hereafter, the average over  $t$  of any variable  $x$  will be denoted by  $\bar{x}$ . Since  $\sum_{t=1}^T (\eta_{it} - \bar{\eta}_i) = 0$ , the second term in the above display disappears and therefore:

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1} - \bar{\tilde{Z}}_i \right)' (\eta_{it} - \bar{\eta}_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \begin{pmatrix} Y_{i,t-1} \\ W_i Y_{t-1} \\ Z'_{i,t-1} \end{pmatrix} (\eta_{it} - \bar{\eta}_i).$$

The  $S_{NT}^*$  is obtained by replacing  $\tilde{Z}_{i,t-1}$  with  $u_{i,t-1}^*$  in the above expression. Because  $u_{i,t-1}^*$  is defined by replacing  $Y_{i,t-1}$  with  $u_{i,t-1}^{**}$  and  $W_i Y_{t-1}$  with  $W_i u_{t-1}^{**}$ , but leaving  $Z_{i,t-1}$  intact, I only need to show that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \begin{pmatrix} Y_{i,t-1} \\ W_i Y_{t-1} \end{pmatrix} (\eta_{it} - \bar{\eta}_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \begin{pmatrix} u_{i,t-1}^{**} \\ W_i u_{t-1}^{**} \end{pmatrix} (\eta_{it} - \bar{\eta}_i) + o_p(1).$$

**a) Time-Lagged Dependent Variable.** First, I concentrate on the component of the score function related to the time-lagged dependent variable and I show that the difference

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (Y_{i,t-1}) (\eta_{it} - \bar{\eta}_i) - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (u_{i,t-1}^{**}) (\eta_{it} - \bar{\eta}_i), \quad (\text{A.5})$$

is  $o_p(1)$ . To establish that (A.5) is  $o_p(1)$ , I rewrite expression (A.5) as

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} (Y_0 - u_0^{**})] (\eta_{it} - \bar{\eta}_i) \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} Y_0] (\eta_{it} - \bar{\eta}_i) \end{aligned} \quad (\text{A.6a})$$

$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} u_0^{**}] (\eta_{it} - \bar{\eta}_i). \quad (\text{A.6b})$$

because  $Y_{i,t-1} - u_{i,t-1}^{**} = \Psi_i^{t-1} [Y_0 - u_0^{**}]$  ( $\Psi_i^{t-1}$  is the  $i^{th}$  row of  $\Psi^{t-1}$ ). Next, I show that (A.6a) and (A.6b) are  $o_p(1)$  because they have a finite mean and a variance of order  $o(1)$ .

**a1)** I first analyze (A.6a). By assumption its expectation is zero because  $Y_0$  is a fixed number and the expectation of  $\eta_{it}$  is by definition zero. To obtain the variance of (A.6a), note that,

$$\mathbf{var} \left[ \sum_{t=1}^T [\Psi_i^{t-1} Y_0] (\eta_{it} - \bar{\eta}_i) \right] = \sum_{t=1}^T [\Psi_i^{t-1} Y_0]^2 \mathbf{var} (\eta_{it} - \bar{\eta}_i) = \mathbf{var} \left[ \eta_{it} - \frac{1}{T} \sum_{s=1}^T \eta_{is} \right] \sum_{t=1}^T [\Psi_i^{t-1} Y_0]^2.$$

Because

$$\mathbf{var} \left[ \eta_{it} - \frac{1}{T} \sum_{s=1}^T \eta_{is} \right] = \frac{T-1}{T} \sigma_\eta^2, \quad \sigma_\eta^2 = E(\eta_{it}^2),$$

the variance of  $\left[ \sum_{t=1}^T [\Psi_i^{t-1} Y_0] (\eta_{it} - \bar{\eta}_i) \right]$  is given by

$$\left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{t=1}^T [\Psi_i^{t-1} Y_0]^2. \quad (\text{A.7})$$

Using (A.7), we can calculate the variance of (A.6a):

$$\mathbf{var} \left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} Y_0]^2 (\eta_{it} - \bar{\eta}_i) \right] = \frac{1}{NT} \left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} Y_0 Y_0' (\Psi_i^{t-1})'] .$$

To obtain an upper bound for the above expression, replace the  $[\Psi_i^{t-1} Y_0 Y_0' (\Psi_i^{t-1})']$  term with its norm. Because the norm of  $[\Psi_i^{t-1} Y_0 Y_0' (\Psi_i^{t-1})']$  is bounded by  $\|\Psi^{t-1} (\Psi^{t-1})'\| \|Y_0\|^2$ , the variance of (A.6a) is no greater than

$$\frac{1}{NT} \left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{i=1}^N \sum_{t=1}^T \|\Psi^{t-1} (\Psi^{t-1})'\| \|Y_0\|^2. \quad (\text{A.8})$$

Furthermore, result (A.1) dictates that  $\|\Psi^{t-1} (\Psi^{t-1})'\| = (k_{\Psi\Psi})^{t-1} < 1$ , and therefore (A.8) becomes

$$\frac{T-1}{NT^2} \sigma_\eta^2 \|Y_0\|^2 \sum_{i=1}^N \sum_{t=1}^T [k_{\Psi\Psi}]^{(t-1)}. \quad (\text{A.9})$$

To simplify expression in (A.9), replace the double sum  $\sum_{i=1}^N \sum_{t=1}^T [k_{\Psi\Psi}]^{t-1}$  with

$$N \left[ \frac{1 - [k_{\Psi\Psi}]^{T-1}}{1 - k_{\Psi\Psi}} \right],$$

to obtain

$$\begin{aligned} &= \frac{T-1}{NT^2} \sigma_\eta^2 \|Y_0\|^2 N \left[ \frac{1 - [k_{\Psi\Psi}]^{T-1}}{1 - k_{\Psi\Psi}} \right] \\ &= \frac{1}{\sqrt{N}} \frac{N}{T} \frac{T-1}{T} \sigma_\eta^2 \frac{\|Y_0\|^2}{\sqrt{N}} \left[ \frac{1 - [k_{\Psi\Psi}]^{T-1}}{1 - k_{\Psi\Psi}} \right]^2. \end{aligned} \quad (\text{A.10})$$

Under condition 1(5),  $\|Y_0\|^2 / \sqrt{N}$  is  $O(1)$ ; under condition 1(2),  $N/T$  is  $O(1)$ ; and under condition

1.3,  $[k_{\Psi\Psi}]^{T-1}$  is  $o(1)$ . Therefore, it follows that (A.10) is  $o(1)$ :

$$\frac{1}{\sqrt{NT}} O(1) \left[ \frac{1 - o(1)}{1 - k_{\Psi\Psi}} \right]^2 = o(1).$$

**a2)** Second, I analyze the term (A.6b) and show that it has a finite mean and a variance of order  $o(1)$ . By assumption, its expectation is zero because  $u_0^{**}$  and  $\eta_t$  are independent. Also, its variance is given by

$$\begin{aligned} \text{var} \left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} u_0^{**}] (\eta_{it} - \bar{\eta}_i) \right] &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} (\Psi_i^{t-1})'] \sigma_{u_0^{**}}^2 \text{var}(\eta_{it} - \bar{\eta}_i) \\ &= \frac{1}{NT} \sigma_{u_0^{**}}^2 \left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{i=1}^N \sum_{t=1}^T [\Psi_i^{t-1} (\Psi_i^{t-1})'], \end{aligned} \quad (\text{A.11})$$

where  $\sigma_{u_0^{**}}^2$  is the variance of the initial conditions for  $u_t^{**}$ . Because the term  $[\Psi_i^{t-1} (\Psi_i^{t-1})']$  is no greater than the norm of  $\Psi^{t-1} (\Psi^{t-1})'$ , expression (A.11) is bounded by

$$\frac{1}{NT} \sigma_{u_0^{**}}^2 \left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{i=1}^N \sum_{t=1}^T \left\| \Psi^{t-1} (\Psi^{t-1})' \right\|. \quad (\text{A.12})$$

By result (A.1), we know that the norm of  $\left\| \Psi^{t-1} (\Psi^{t-1})' \right\| (= k_{\Psi\Psi})$  is less than one, and therefore the expression in (A.12) can be written as:

$$\begin{aligned} \frac{1}{NT} \sigma_{u_0^{**}}^2 \left( \frac{T-1}{T} \sigma_\eta^2 \right) \sum_{i=1}^N \sum_{t=1}^T [k_{\Psi\Psi}]^{(t-1)} &= , \\ &= \frac{\sigma_{u_0^{**}}^2}{NT} \left( \frac{T-1}{T} \sigma_\eta^2 \right) N \left[ \frac{1 - [k_{\Psi\Psi}]^{T-1}}{1 - k_{\Psi\Psi}} \right]^2 = \frac{1}{T} O(1) \left[ \frac{1 - o(1)}{1 - k_{\Psi\Psi}} \right]^2 = o(1). \end{aligned}$$

**b) Spatially Lagged Dependent Variable.** Next, I move to the part of the score function that is related to the spatially lagged dependent variable and I show that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (W_i Y_{t-1}) (\eta_{it} - \bar{\eta}_i) - \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (W_i u_{i,t-1}^{**}) (\eta_{it} - \bar{\eta}_i) = o_p(1).$$

Using  $W_i (Y_{t-1} - u_{t-1}^{**}) = \tilde{\Psi}_i^{t-1} [Y_0 - u_0^{**}]$ ,  $\tilde{\Psi}_i = W_i \Psi$ , I rewrite the left side of the above expression as follows:

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left[ \tilde{\Psi}_i^{t-1} (Y_0 - u_0^{**}) \right] (\eta_{it} - \bar{\eta}_i)$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left[ \tilde{\Psi}_i^{t-1} Y_0 \right] (\eta_{it} - \bar{\eta}_i) \quad (\text{A.13a})$$

$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left[ \tilde{\Psi}_i^{t-1} u_0^{**} \right] (\eta_{it} - \bar{\eta}_i). \quad (\text{A.13b})$$

Compared to expressions (A.6a) and (A.6b), expressions (A.13a) and (A.13b) replace  $\Psi$  with  $\tilde{\Psi}$ ,  $\tilde{\Psi} = W\Psi$ . Consequently, I repeat the same steps as in the case of the component of the score related to  $Y_{t-1}$ , to prove that the approximation with respect to  $WY_{t-1}$  is of order  $o_p(1)$ .

We can take the same steps as in (b) above because the bound of the norm of  $\tilde{\Psi}$  is the same as the bound of  $\Psi$ :  $\|\tilde{\Psi}\| = \|W\Psi\| \leq \|W\| \|\Psi\| = \|\Psi\|$ . Therefore, one can show that  $\|\tilde{\Psi}^s\| \leq [|\rho_1| + |\pi_1|]^s$ , and using result (A.2), one can obtain that:

$$\left\| \sum_{t=0}^T \tilde{\Psi}^{t-1} \right\| \leq \frac{1 - [|\rho_1| + |\pi_1|]^{T-1}}{1 - |\rho_1| - |\pi_1|},$$

because  $\|W\| = 1$  and .

### A.3 Asymptotic Properties of Estimator

In this part, I find the limiting distributions of  $\hat{\varphi}_b$  (theorem 1), and the limiting distribution of the bias-corrected estimator  $\hat{\varphi}_c$  (theorem 2). To do so, I prove lemmas 4 and 5.

**Lemma 4.** Under conditions 1 to 5,  $S_{NT}^* \rightarrow \mathcal{N}(B, V^*)$ . The vector  $B$  is the limit of  $B_{NT}$ :

$$B_{NT} = \left[ -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \text{tr}\{\Pi\}, \quad -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \text{tr}\{W\Pi\}, \quad -\sqrt{\frac{N}{T}} \frac{T-1}{T} (\sigma_{z\eta}) \right]',$$

where the  $1 \times K$  vector  $\sigma_{x\eta} = (\sigma_{1,x\eta}, \dots, \sigma_{K,x\eta})$  contains the expectations  $\sigma_{k,x\eta} = \mathbf{E}(X_{k,it}\eta_{it})$ ,  $k = \{1, \dots, K\}$ . The  $1 \times K$  vector  $\sigma_{z\eta} = (\sigma_{1,z\eta}, \dots, \sigma_{K,z\eta})$  contains the expectations  $\sigma_{k,z\eta} = \mathbf{E}(Z_{k,it}\eta_{it})$ ,  $k = \{1, \dots, K\}$  ( $Z_k$  is the instrument for the  $k^{th}$  control variable  $X_k$ ). The matrix  $\Pi$  is

$$\Pi = \frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1} = \frac{T-1}{T} (I - \Psi)^{-1} - \frac{1}{T} \Psi (I - \Psi)^{-1} (I - \Psi^{T-1}) = (I - \Psi)^{-1} + o(1),$$

where  $\Psi = (\pi_1 I + \rho_1 W)$ , and  $\text{tr}$  is the trace operator. Also,  $V^*$  is the asymptotic variance of  $S_{NT}^*$ .

**Proof for Lemma 4.** First, I write the approximate score function  $S_{NT}^*$  as:

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \begin{bmatrix} u_{i,t-1}^{**} \\ W_i u_{i,t-1}^{**} \\ Z_{i,t-1}' \end{bmatrix} (\eta_{it} - \bar{\eta}_i) = \frac{1}{\sqrt{NT}} \sum_{t=1}^T \begin{bmatrix} (u_{t-1}^{**})' \\ (u_{t-1}^{**})' W' \\ Z_{t-1}' \end{bmatrix} (\eta_t - \bar{\eta}). \quad (\text{A.14})$$

To prove the lemma, I follow the same approach as in Hahn and Kuersteiner (2002) and a) I show that the expectation of (A.14) is bounded at the limit, b) I show that the  $(NT)^{-1/2} \sum_{i,t} u_{i,t-1}^* \eta_{it}$  is normally distributed, and c) I define the asymptotic variance of the estimator.

### a) Expectation

The approximate score function  $S_{NT}^*$  has  $K + 2$  components, and in what follows I calculate their expectations and show that they are bounded.

**a1)** I calculate the expectation of  $(1/\sqrt{NT}) \sum_{t=1}^T (u_{t-1}^{**})' (\eta_t - \bar{\eta})$  and its bound in six steps.

**Step 1:** Because  $u_{t-1}^{**}$  is independent of  $\eta_{t+s}$ ,  $s \geq 0$ ,

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' (\eta_t - \bar{\eta}) \right] &= - \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' \frac{1}{T} \sum_{s=1}^T \eta_s \right] \\ &= - \frac{1}{\sqrt{NT}} \sum_{t=2}^T \sum_{s=1}^{t-1} \frac{1}{T} \mathbf{E} \left[ (u_{t-1}^{**})' \eta_s \right] = - \frac{1}{\sqrt{NT}} \sum_{t=2}^T \sum_{s=1}^{t-1} \text{tr} \left\{ \frac{1}{T} \mathbf{E} \left[ (u_{t-1}^{**})' \eta_s \right] \right\} \\ &= - \frac{1}{\sqrt{NT}} \sum_{t=2}^T \sum_{s=1}^{t-1} \text{tr} \left\{ \frac{1}{T} \mathbf{E} \left[ \eta_s (u_{t-1}^{**})' \right] \right\} \end{aligned} \quad (\text{A.15})$$

The above expression contains the expectation  $\mathbf{E} \left[ \eta_s (u_{t-1}^{**})' \right]$ , which I compute in step 2.

**Step 2:** To calculate the expectation  $\mathbf{E} \left[ \eta_s (u_{t-1}^{**})' \right]$  in (A.15), I use the fact that

$$u_{t-1}^{**} = \Psi^{t-1} u_0^{**} + \sum_{s=1}^{t-1} \Psi^{s-1} (c + X_{t-s} \lambda + \eta_{t-s}),$$

which implies that  $\text{tr} \left\{ \mathbf{E} \left[ \eta_s (u_{t-1}^{**})' \right] \right\} = (\sigma_\eta^2 + \sigma_{x\eta} \lambda) \text{tr}((\Psi^{s-1})') = (\sigma_\eta^2 + \sigma_{x\eta} \lambda) \text{tr}(\Psi^{s-1})$ . Therefore,

$$\mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' (\eta_t - \bar{\eta}) \right] = - \frac{1}{\sqrt{NT}} (\sigma_\eta^2 + \sigma_{x\eta} \lambda) \text{tr} \left\{ \frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1} \right\}, \quad (\text{A.16})$$

where  $\Psi = (\pi_1 I + \rho_1 W)$ .

**Step 3:** By the usual result on Cesaro averages, I simplify the double sum in the right-hand-side of (A.16) and I bound it with  $(I - \Psi)^{-1}$  plus an  $o(1)$  term:

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1} &= \frac{1}{T} \sum_{t=2}^T (I - \Psi)^{-1} (I - \Psi^{t-1}) = \frac{T-1}{T} (I - \Psi)^{-1} - \frac{1}{T} \Psi \sum_{t=2}^T \Psi^{t-2} \\ &= \frac{T-1}{T} (I - \Psi)^{-1} - \frac{1}{T} \Psi (I - \Psi)^{-1} (I - \Psi^{T-1}) \\ &= (I - \Psi)^{-1} + o(1). \end{aligned} \quad (\text{A.17})$$

**Step 4:** I use the result in (A.17) to replace the double sum  $\frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1}$  in the right-hand-side of (A.16) with  $[(I - \Psi)^{-1} + o(1)]$ . Therefore, I obtain the following expression for the expectation

of  $(1/\sqrt{NT})\sum_{t=1}^T (u_{t-1}^{**})' (\eta_t - \bar{\eta})$ :

$$\mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' (\eta_{it} - \bar{\eta}_i) \right] = - \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr} \{ (I - \Psi)^{-1} + o(1) \}. \quad (\text{A.18})$$

**Step 5:** I show that the expectation (A.18) is bounded under the HK asymptotics in condition 1(2). To establish this result, I rewrite the right-hand-side of (A.18) by replacing  $(I - \Psi)^{-1}$  with

$$\frac{1}{1 - \pi_1} \left[ I - \frac{\rho_1}{1 - \pi_1} W \right]^{-1}$$

and obtain the following expression:

$$- \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr} \{ (I - \Psi)^{-1} \} = - \frac{1}{\sqrt{NT}} \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{1 - \pi_1} \mathbf{tr} \left\{ \left[ I - \frac{\rho_1}{1 - \pi_1} W \right]^{-1} \right\}.$$

**Step 6:** Because the norm of  $[I - [\rho_1/(1 - \pi_1)]W]^{-1}$  is bounded by  $\frac{1}{1 - |\rho_1/(1 - \pi_1)|}$  (see Lemma 1), the trace of the matrix  $[I - [\rho_1/(1 - \pi_1)]W]^{-1}$  is bounded by  $\frac{1}{1 - |\rho_1/(1 - \pi_1)|}N$ . Therefore:

$$\mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' (\eta_{it} - \bar{\eta}_i) \right] \leq - \frac{1}{\sqrt{NT}} \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{1 - \pi_1} \frac{N}{1 - |\rho_1/(1 - \pi_1)|}. \quad (\text{A.19})$$

The right-hand-side of (A.19) equals

$$- \sqrt{\frac{N}{T}} \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{1 - \pi_1} \frac{1}{1 - |\rho_1/(1 - \pi_1)|},$$

and its  $O(1)$  because  $\sqrt{N/T}$  is bounded under condition 1(2). Thus, the expectation of

$$(1/\sqrt{NT}) \sum_{t=1}^T (u_{t-1}^{**})' (\eta_{it} - \bar{\eta}_i)$$

is bounded.

**a2)** I calculate the expectation of  $(1/\sqrt{NT})\sum_{t=1}^T (u_{t-1}^{**})' W' (\eta_t - \bar{\eta})$  and its bound in six steps, similar to the analysis in (a1) above.

**Step 1:** Because  $u_{t-1}^{**}$  is independent of  $\eta_{t+s}$ ,  $s \geq 0$ ,

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' W' (\eta_t - \bar{\eta}) \right] &= - \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' W' \frac{1}{T} \sum_{s=1}^T \eta_s \right] \\ &= - \frac{1}{\sqrt{NT}} \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbf{tr} \left\{ \frac{1}{T} \mathbf{E} \left[ (u_{t-1}^{**})' W' \eta_s \right] \right\} = - \frac{1}{\sqrt{NT}} \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbf{tr} \left\{ \frac{1}{T} \mathbf{E} \left[ W' \eta_s (u_{t-1}^{**})' \right] \right\}. \end{aligned}$$



**Step 2:** Because the trace of the expectations  $\mathbf{E} \left[ W' \eta_s (u_{t-1}^{**})' \right]$  is given by  $(\sigma_\eta^2 + \sigma_{x\eta}\lambda) \mathbf{tr}(W'(\Psi^{s-1})') = (\sigma_\eta^2 + \sigma_{x\eta}\lambda) \mathbf{tr}(W\Psi^{s-1})$ ,

$$\begin{aligned} & \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' W' (\eta_t - \bar{\eta}) \right] \\ &= -\frac{1}{\sqrt{NT}} (\sigma_\eta^2 + \sigma_{x\eta}\lambda) \sum_{t=2}^T \sum_{s=1}^{t-1} \mathbf{tr} \left\{ \frac{1}{T} W \Psi^{s-1} \right\} \\ &= -\frac{1}{\sqrt{NT}} (\sigma_\eta^2 + \sigma_{x\eta}\lambda) \mathbf{tr} \left\{ W \frac{1}{T} \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1} \right\}, \quad \Psi = \pi_1 I + \rho_1 W. \end{aligned}$$

**Step 3:** Note that the double sum  $(1/T) \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1}$  equals  $(I - \Psi)^{-1}$  plus an  $o(1)$  term.

**Step 4:** I replace  $(1/T) \sum_{t=2}^T \sum_{s=1}^{t-1} \Psi^{s-1}$  with  $(I - \Psi)^{-1}$  and I obtain the following expression for the expectation of  $(1/\sqrt{NT}) \sum_{t=1}^T (u_{t-1}^{**})' W' (\eta_t - \bar{\eta})$ :

$$\mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T (u_{t-1}^{**})' W' (\eta_t - \bar{\eta}) \right] = -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr} \{ W (I - \Psi)^{-1} + o(1) \}. \quad (\text{A.20})$$

**Step 5:** To show that (A.20) is bounded, I replace  $(I - \Psi)^{-1}$  with  $\frac{1}{1-\pi_1} [I - [\rho_1/(1-\pi_1)] W]^{-1}$ :

$$-\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \mathbf{tr} \{ W' (I - \Psi)^{-1} \} = -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \frac{1}{1-\pi_1} \mathbf{tr} \left\{ W \left[ I - \frac{\rho_1}{1-\pi_1} W \right]^{-1} \right\}. \quad (\text{A.21})$$

**Step 6:** Note that the norm of the matrix  $[I - [\rho_1/(1-\pi_1)] W]^{-1}$  is  $\frac{1}{1-|\rho_1/(1-\pi_1)|}$  (see Lemma 1). I therefore bound the matrix  $[I - [\rho_1/(1-\pi_1)] W]^{-1}$  in (A.21) with

$$\frac{1}{1-|\rho_1/(1-\pi_1)|} \mathbf{1}_{N \times N}$$

( $\mathbf{1}_{N \times N}$  is an  $N$  by  $N$  matrix of ones). The previous step implies that the right side of (A.21) is no greater than

$$\begin{aligned} & -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \frac{1}{1-\pi_1} \mathbf{tr} \left\{ W \frac{1}{1-|\rho_1/(1-\pi_1)|} \mathbf{1}_{N \times N} \right\} \\ &= -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \frac{1}{1-\pi_1} \frac{1}{1-|\rho_1/(1-\pi_1)|} \mathbf{tr} \{ W \mathbf{1}_{N \times N} \}. \end{aligned} \quad (\text{A.22})$$

Because the rows of  $W$  sum to 1,  $\mathbf{tr} \{ W \mathbf{1}_{N \times N} \} = \mathbf{tr} \{ \mathbf{1}_{N \times N} \} = N$ . Thus, the expression (A.22) equals:

$$= -\frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \frac{1}{1-\pi_1} \frac{N}{1-|\rho_1/(1-\pi_1)|} = -\sqrt{\frac{N}{T}} \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{1-\pi_1} \frac{1}{1-|\rho_1/(1-\pi_1)|}.$$

The above expression is  $O(1)$  because  $\sqrt{N/T}$  is bounded under condition 1(2).

**a3)** Third, I compute the expectation of  $\frac{1}{\sqrt{NT}} \sum_{t=1}^T (Z'_{t-1}) (\eta_t - \bar{\eta})$ ,  $Z_t : N \times K$ ,  $\eta_t : N \times 1$ . To do so, I focus on the expectation related to the  $k^{th}$  explanatory variable:

$$\begin{aligned} \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^T Z'_{k,t-1} (\eta_t - \bar{\eta}) \right] &= - \mathbf{E} \left[ \frac{1}{\sqrt{NT}} \frac{1}{T} \sum_{t=1}^T Z'_{k,t-1} \sum_{s=1}^T \eta_s \right] \\ &= - \frac{1}{\sqrt{NT}} \frac{1}{T} \sum_{t=2}^T \text{tr} \{ \mathbf{E} [Z'_{k,t-1} \eta_{t-1}] \} = - \frac{1}{\sqrt{NT}} \frac{1}{T} \sum_{t=2}^T \text{tr} \{ \mathbf{E} [\eta_{t-1} Z'_{k,t-1}] \} \\ &= - \frac{1}{\sqrt{NT}} \frac{T-1}{T} (\sigma_{k,z\eta}) \text{tr} \{ I \} = - \sqrt{\frac{N}{T}} \frac{T-1}{T} \sigma_{k,z\eta}, \end{aligned}$$

where  $\sigma_{k,z\eta}$  is  $\mathbf{E}(X_{z,it} \eta_{it})$ . Under condition 1(2),  $\sqrt{N/T}$  is always finite, and therefore the above expectation is bounded.

### b) Asymptotic Normality.

I now show that the double sum

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{i=1}^N u_{i,t-1}^* \eta_{it}, \quad (u_{i,t-1}^*)' = (u_{i,t-1}^{**}, W_i u_{i,t-1}^{**}, Z_{i,t-1}),$$

converges to a normal distribution. I prove asymptotic normality using a central limit theorem (CLT) for martingale difference sequence (mds). This approach is similar to that of Driscoll and Kraay (1998).

I apply the CLT on the variable  $\xi_{N,t}$ ,

$$\xi_{N,t} = \frac{1}{\sqrt{N}} \sum_{i=1}^N u_{i,t-1}^* \eta_{it}.$$

The process  $\xi_{N,t}$  is a martingale difference sequence because  $\mathbf{E} [\xi_{N,t} | \mathcal{F}_{t-1}] = 0$ , where  $\mathcal{F}_{t-1}$  is the sigma algebra generated by all variables dated  $t-1$  or earlier. Under this setup, the double sum

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{i=1}^N u_{i,t-1}^* \eta_{it} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \xi_{N,t}$$

converges to a normal distribution by the CLT for mds in Corollary 5.25 by White (1994). The CLT operates under three conditions:  $\mathbf{E} [\xi_{N,t} \xi'_{N,t}]$  is finite,  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  is finite for some  $\zeta > 0$  and all  $t$ , and the probability limit of  $T^{-1} \sum_t \xi_{N,t} \xi'_{N,t}$  is  $\sigma_\xi^2$  and  $\sigma_\xi^2$  is finite. Next, I prove the three aforementioned conditions.

**b1)**  $\mathbf{E} [\xi_{N,t} \xi'_{N,t}]$  is finite:

$$\begin{aligned} \mathbf{E} [\xi_{N,t} \xi'_{N,t}] &= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \mathbf{E} [u_{i,t-1}^* \eta_{it} (u_{j,t-1}^*)' \eta_{jt}], \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{E} [u_{i,t-1}^* (u_{i,t-1}^*)'] \mathbf{E} \eta_{it}^2 + \frac{1}{N} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \mathbf{E} [u_{i,t-1}^* (u_{j,t-1}^*)'] \mathbf{E} \eta_{it} \mathbf{E} \eta_{jt}. \end{aligned} \tag{A.24}$$

Because  $\mathbf{E}\eta_{it} = \mathbf{E}\eta_{jt} = 0$ , the above expression simplifies to:

$$\begin{aligned} &= \frac{\sigma_\eta^2}{N} \sum_{i=1}^N \mathbf{E} \left[ u_{i,t-1}^* (u_{i,t-1}^*)' \right], \quad (u_{i,t-1}^*)' = (u_{i,t-1}^{**}, W_i u_{i,t-1}^{**}, Z_{i,t-1}), \\ &= \frac{\sigma_\eta^2}{N} \sum_{i=1}^N \mathbf{E} \begin{bmatrix} (u_{i,t-1}^{**})^2 & u_{i,t-1}^{**} W_i u_{i,t-1}^{**} & u_{i,t-1}^{**} Z_{i,t-1} \\ u_{i,t-1}^{**} W_i u_{i,t-1}^{**} & (W_i u_{i,t-1}^{**})^2 & W_i u_{i,t-1}^{**} Z_{i,t-1} \\ Z_{i,t-1}' u_{i,t-1}^{**} & Z_{i,t-1}' W_i u_{i,t-1}^{**} & Z_{i,t-1}' Z_{i,t-1} \end{bmatrix}. \end{aligned} \quad (\text{A.25})$$

Next, I show that all the elements of  $\mathbf{E} \left[ u_{i,t-1}^* (u_{i,t-1}^*)' \right]$  are finite:

- By Condition 4, the expectations  $\mathbf{E} \left[ Z_{i,t-1}' Z_{i,t-1} \right]$  and  $\mathbf{E} \left[ Z_{i,t-1}' u_{i,t-1}^{**} \right]$  are finite.
- The expectation of  $(u_{i,t-1}^{**})^2$  is  $O(1)$  because  $\mathbf{E} \left[ u_{i,t-1}^{**} (u_{i,t-1}^{**})' \right]$  is  $O(1)$ :

$$\begin{aligned} &\mathbf{E} \left[ u_{i,t-1}^{**} (u_{i,t-1}^{**})' \right] \quad (\text{A.26}) \\ &= \mathbf{E} \left[ \sum_{s=0}^{\infty} \Psi^s (c + X_{t-s} \lambda + \eta_{t-s}) (c' + \lambda' X_{t-s}' + \eta_{t-s}') (\Psi^s)' \right] \\ &= \sum_{s=0}^{\infty} \Psi^s H (\Psi^s)', \quad H = (cc' + c\lambda' \mathbf{E}(X_t') + \mathbf{E}(X_t) \lambda c' + \mathbf{E}(X_t \lambda \lambda' X_t') + 2\sigma_{x\eta} \lambda' I + \sigma_\eta^2 I), \\ &\leq \sum_{s=0}^{\infty} \|\Psi^s\| \|H\| \|(\Psi^s)'\| = \|H\| \sum_{s=0}^{\infty} (\|\Psi\| \|\Psi'\|)^s \\ &= \frac{\|H\|}{1 - k_{\Psi\Psi}} = O(1). \end{aligned}$$

The last equality in the above expression follows from the fact that the product  $\|\Psi\| \|\Psi'\|$  is given by  $k_{\Psi\Psi}$ , which is less than one. See Results 1 and 2 in Section A.1.

- The  $\mathbf{E} \left[ u_{i,t-1}^{**} W_i u_{i,t-1}^{**} \right]$  is  $O(1)$  because it is a weighted average of the  $i^{th}$  row of  $\mathbf{E} \left[ u_{i,t-1}^{**} (u_{i,t-1}^{**})' \right]$ , which is finite (see expression (A.26) above):

$$\mathbf{E} \left[ u_{i,t-1}^{**} W_i u_{i,t-1}^{**} \right] = \sum_{j=1}^N w_{ij} \mathbf{E} \left[ u_{i,t-1}^{**} u_{j,t-1}^{**} \right] = \sum_{j=1}^N w_{ij} O(1) = O(1). \quad (\text{A.27})$$

The last equality in the above display follows from the fact that the rows of  $W$  always sum to 1 (see condition 2(3)).

- The  $\mathbf{E} (W_i u_{t-1}^{**})^2$  is also  $O(1)$ :

$$\begin{aligned} \mathbf{E} \left[ (W_i u_{t-1}^{**})^2 \right] &= \mathbf{E} \left( \sum_{j=1}^N w_{ij} u_{j,t-1}^{**} \right)^2 = \sum_{j=1}^N w_{ij} \sum_{l=1}^N w_{il} [\mathbf{E} u_{j,t-1}^{**} u_{l,t-1}^{**}] \\ &= \sum_{j=1}^N w_{ij} \sum_{l=1}^N w_{il} O(1) = \sum_{j=1}^N w_{ij} O(1) = O(1). \end{aligned} \quad (\text{A.28})$$

The calculations in the above display use the fact that the rows of  $W$  always sum to 1 and that the  $\mathbf{E} \left[ u_{t-1}^{**} (u_{t-1}^{**})' \right]$  is finite (see expression (A.26) above).

- Finally, using similar steps as in (A.27), the expectation of  $Z_{i,t-1} W_i u_{t-1}^{**}$  can be shown to be  $O(1)$ .

Therefore, the  $\mathbf{E} [\xi_{N,t} \xi'_{N,t}]$  is  $O(1)$ .

**b2)** The expectation  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  is a vector with  $K+2$  elements. Next, I show that each of these elements is finite.

- The first element of  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  is

$$\mathbf{E} \left| \sum_{t=1}^T u_{i,t-1}^{**} \eta_{it} \right|^{2+\zeta} \leq \mathbf{E} \left[ \sum_{t=1}^T |u_{i,t-1}^{**} \eta_{it}| \right]^{2+\zeta}.$$

Because by (A.3),  $u_{i,t-1}^{**} = \Psi_i^{t-1} u_0^{**} + \sum_{s=0}^{t-2} \Psi_i^s (c + X_{t-1-s} \lambda + \eta_{t-1-s})$ , the above expression equals:

$$\mathbf{E} \left[ \sum_{t=1}^T \left| \Psi_i^{t-1} u_0^{**} \eta_{it} + \sum_{s=0}^{t-2} \Psi_i^s (c + X_{t-1-s} \lambda + \eta_{t-1-s}) \eta_{it} \right| \right]^{2+\zeta}. \quad (\text{A.29})$$

Because the absolute value of

$$\Psi_i^{t-1} u_0^{**} \eta_{it} + \sum_{s=0}^{t-2} \Psi_i^s (c + X_{t-1-s} \lambda + \eta_{t-1-s}) \eta_{it}$$

is bounded by the sum of the absolute value of its components, expression (finite) is bounded by

$$\begin{aligned} &\mathbf{E} \left[ \sum_{t=1}^T \left| \Psi_i^{t-1} u_0^{**} \eta_{it} \right| + \sum_{t=1}^T \sum_{s=0}^{t-2} \left| \Psi_i^s (c + X_{t-1-s} \lambda + \eta_{t-1-s}) \eta_{it} \right| \right]^{2+\zeta} \\ &\leq \mathbf{E} \left[ \sum_{t=1}^T \sum_{j=1}^N \left| \Psi_{ij}^{t-1} \right| |u_{0,j}^{**}| |\eta_{it}| + \sum_{t=1}^T \sum_{s=0}^{t-2} \sum_{j=1}^N \left| \Psi_{ij}^s \right| \{ |c_j| + |X_{j,t-1-s}| |\lambda| + |\eta_{j,t-1-s}| \} |\eta_{it}| \right]^{2+\zeta} \end{aligned} \quad (\text{A.30})$$

where  $\Psi_{ij}^t$  is the  $ij^{th}$  element of  $\Psi^t$  ( $\Psi_{ij}^t$  is finite because  $\Psi^t$  is absolutely summable (see lemma 2)). The term, which is raised to the power of  $(2+\zeta)$ , inside the expectation in the above

display can be expanded using the multinomial theorem. The expectation of this expansion is finite because it includes scaled cross-products of the expectations of  $\mathbf{E} |\eta_{it}|^\delta$ ,  $\sum_{k=1}^K \mathbf{E} |X_{k,it}|^\delta$ ,  $\sum_{k=1}^K \mathbf{E} |X_{k,it} \eta_{it}|^\delta$ , and  $\mathbf{E} |u_{0,i}^{**}|^\delta$ ,  $0 < \delta \leq 2 + \zeta$ , which are finite by assumption (see Conditions 1(1), 3 and 5). The scaling factors are  $\Psi_{ij}^t$ ,  $\lambda$  and  $c_i$ . Thus,  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  is bounded.

- The second element of  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  is  $\mathbf{E} \left[ \sum_{t=1}^T |W_i u_{t-1}^{**} \eta_{it}| \right]^{2+\zeta}$ . To show that it is finite, repeat the steps in (A.29) and (A.30) by replacing the  $ij^{th}$  element of  $\Psi^t$  with the  $ij^{th}$  element of  $W\Psi^t$ .
- The last  $K$  elements of  $\mathbf{E} |\xi_{N,t}|^{2+\zeta}$  are  $\mathbf{E} \left[ \sum_{t=1}^T |Z_{k,i,t-1} \eta_{it}| \right]^{2+\zeta}$ ,  $k = \{1, \dots, K\}$ , which are finite by assumption (see Conditions 1(1) and 4).

**b3)** The  $\sigma_\xi^2$  is finite because the expectation of  $T^{-1} \sum_t \xi_{N,t} \xi'_{N,t}$  is  $O(1)$ :

$$\frac{1}{T} \sum_{t=1}^T \mathbf{E} [\xi_{N,t} \xi'_{N,t}] = \frac{1}{TN} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N \mathbf{E} \left[ u_{i,t-1}^* \eta_{it} (u_{j,t-1}^*)' \eta_{jt} \right].$$

As in expressions (A.24) and (A.25), the above expression equals to

$$\frac{1}{T} \sum_{t=1}^T \mathbf{E} [\xi_{N,t} \xi'_{N,t}] = \frac{1}{TN} \sigma_\eta^2 \sum_{t=1}^T \sum_{i=1}^N \mathbf{E} \left[ u_{i,t-1}^* (u_{i,t-1}^*)' \right].$$

because  $\mathbf{E} \eta_{it} = \mathbf{E} \eta_{jt} = 0$ . I have already shown in part (a) above that  $\mathbf{E} \left[ u_{i,t-1}^* (u_{i,t-1}^*)' \right]$  is  $O(1)$  (see expressions (A.26) to (A.28)), which implies that

$$\frac{1}{T} \sum_{t=1}^T \mathbf{E} [\xi_{N,t} \xi'_{N,t}] = \frac{1}{TN} \sigma_\eta^2 \sum_{t=1}^T \sum_{i=1}^N O(1) = O(1).$$

### c) Variance.

I have shown that the expectation of (A.14) is bounded at the limit, and that  $\xi_{N,t}/\sqrt{T}$  converges to a normal distribution. I now define the variance-covariance matrix of (A.14). The variance-covariance matrix of  $S_{NT}^*$  is the probability limit of  $\frac{1}{NT} [S_{NT}^* - B_{NT}] [S_{NT}^* - B_{NT}]'$ , which is denoted by  $V^*$ .

**Lemma 5.** Under conditions 1 to 5, the sample score function,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' (\eta_{it} - \bar{\eta}_i),$$

converges to  $N(B, V^*)$ .

**Proof for Lemma 5.** The result follows from lemmas 3 and 4.

**Theorem 1.** Under conditions 1 to 5,  $\sqrt{NT}(\hat{\varphi}_b - \varphi) \rightarrow \mathcal{N}(B, Q^{-1}VQ^{-1})$ , where

$$Q = \text{plim} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \left( \tilde{X}_{it}^d \right), \quad V = \text{plim} \frac{1}{NT} [\bar{S} - \mathbf{E}\bar{S}] [\bar{S} - \mathbf{E}\bar{S}]', \text{ and}$$

$$\bar{S} = \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' (\eta_t - \bar{\eta}).$$

**Proof for Theorem 1.** First note that  $V$  equals  $V^*$  because  $S_{NT} = S_{NT}^* + o_p(1)$ . Then, the proof follows from lemma 5.

**Theorem 2.** Under conditions 1 to 5,  $\sqrt{NT}(\hat{\varphi}_c - \varphi) \rightarrow \mathcal{N}(0, Q^{-1}VQ^{-1})$ .

**Proof for Theorem 2.** First note that  $V$  is equal to  $V^*$  because  $S_{NT} = S_{NT}^* + o_p(1)$ . Second,  $\hat{B}_{NT}$  is the value of  $B_{NT}$  under a consistent estimator of  $\varphi$ ,  $\hat{\varphi}_0$ . Because  $\hat{\varphi}_0 = \varphi + o_p(1)$ ,  $\hat{B}_{NT}$  converges to  $B$ . Therefore, lemma 5 implies that

$$\left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' (\eta_{it}) - \hat{B}_{NT} \right] \rightarrow \mathcal{N}(0, V).$$

#### A.4 Higher-Order Autoregression

In this part, the analysis is generalized to include  $m_1$  time-lagged dependent variables and  $m_2$  spatially lagged dependent variables. Let us call this model the  $AR(m_1, m_2)$  model. In matrix form, the  $AR(m_1, m_2)$  model is

$$Y_t = \sum_{r=1}^{m_1} \pi_r Y_{t-r} + \sum_{r=1}^{m_2} \rho_r W Y_{t-r} + X_t \lambda + \eta_t + c \quad \text{for } t = (1, \dots, T), \quad (\text{A.31})$$

where the  $N \times 1$  vectors  $Y_t$ ,  $\eta_t$ , and  $c$ , stack together the observations of all cross-sectional units at time  $t$ . For example,  $Y_t = (Y_{1t}, \dots, Y_{Nt})'$ . Similarly,  $X_t$  is an  $N \times K$  matrix of explanatory variables:  $X_t = (X_{1,t}, \dots, X_{K,t})$ , where  $X_{k,t}$  is an  $N \times 1$  vector,  $k = \{1, \dots, K\}$ .

The timing convention is that the available data for estimation is from  $t = -m + 1$  to  $t = T$ . Therefore, the  $m$  vectors  $Y_{-m+1}, \dots, Y_0$  are the initial conditions of  $Y$  ( $m = \max(m_1, m_2)$ ). Like Hahn and Kuersteiner (2004), these initial conditions are nonrandom and known to the econometrician.

Similar to the case with one time-lagged and one spatially-lagged dependent variables, the score function of the hybrid estimator is

$$S_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \left( \tilde{Z}_{i,t-1}^d \right)' \eta_{it}^d,$$

where  $\tilde{Z}_{i,t-1}^d$  is  $(Y_{i,t-1}^d, \dots, Y_{i,t-m_1}^d, W_i Y_{i,t-1}^d, \dots, W_i Y_{i,t-m_2}^d, Z_{i,t-1}^d)$ . The approximate score function,  $S_{NT}^*$ , is defined by replacing  $\tilde{Z}_{i,t-1}^d$  with  $u_{i,t-1}^*$ ,

$$(u_{i,t-1}^*)' \equiv (u_{i,t-1}^{**}, \dots, u_{i,t-m_1}^{**}, W_i u_{i,t-1}^{**}, \dots, W_i u_{i,t-m_2}^{**}, Z_{i,t-1})'$$

The variable  $u^{**}$  in  $u^*$  is an approximation to the dependent variable  $Y$ . The approximation assumes

that that, unlike  $Y$ ,  $u^{**}$  is initialized in the infinite past. By collecting the observations of all cross-sectional units at all time periods,  $u^{**}$  is given by

$$u^{**} = (\Gamma_L)^{-1} (\tilde{\eta} + u_a^{**}), \quad (\text{A.32})$$

where  $\tilde{\eta} = (i_T \otimes c + X\lambda + \eta)$ . The random variables  $u^{**}$  and  $\eta$  are  $NT \times 1$  vectors, and the matrix of the control variable  $X$  is an  $NT \times K$  matrix. For example,  $\eta_t = (\eta_{1t}, \dots, \eta_{Nt})'$  and  $\eta = (\eta'_1, \dots, \eta'_T)'$ . Also,  $i_T$  denotes a  $T \times 1$  vector of ones, and  $\otimes$  denotes the Kronecker product. The  $m$  vectors  $u_{-m+1}^{**}, \dots, u_0^{**}$  are the initial conditions of  $u^{**}$ . These initial conditions are independent of  $\eta$ .

The random variable  $u_a^{**}$  in the right side of (A.32) summarizes the past of  $u^{**}$ , and  $u_a^{**}$  is an  $NT \times 1$  vector. Its first  $mT$  rows are  $u_a^{**} = (u_{a1}^{**}, \dots, u_{am}^{**})'$  while the rest of its elements are zero. The variable  $u_{aj}^{**}$  is an  $N \times 1$  vector, which is defined as a linear combination of the initial conditions of  $u^{**}$ :

$$u_{aj}^{**} = \mathbf{I}_{j \leq m_1} \left[ \sum_{r=j}^{m_1} \pi_r u_{j-r}^{**} \right] + \mathbf{I}_{j \leq m_2} \left[ \sum_{r=j}^{m_2} \rho_r W u_{j-r}^{**} \right], \quad (\text{A.33})$$

where  $j$  is  $\{1, \dots, m\}$ , and  $\mathbf{I}$  is the indicator function ( $\mathbf{I}_{j \leq \kappa} = 1$  for  $j \leq \kappa$ ;  $\mathbf{I}_{j \leq \kappa} = 0$  for  $j > \kappa$ ). For example, when  $m_1 = 1$  and  $m_2 = 2$ ,  $u_{a1}^{**} = \pi_1 u_0^{**} + \rho_1 W u_0^{**} + \rho_2 W u_{-1}^{**}$ , and  $u_{a2}^{**} = \rho_2 W u_0^{**}$ . Because the initial condition of  $u^{**}$  are independent of  $\eta$ ,  $u_a^{**}$  is also independent of  $\eta$ .

The block matrix  $\Gamma_L$  in the right side of (A.32) is an  $NT \times NT$  lower Toeplitz matrix of the following form:

$$\Gamma_L = [I_T \otimes I_N] - \sum_{r=1}^{m_1} \pi_r [L^r \otimes I_N] - \sum_{r=1}^{m_2} \rho_r [L^r \otimes W],$$

where  $I_T$  is the  $T \times T$  identity matrix and  $I_N$  is the  $N \times N$  identity matrix. The  $T \times T$  matrix  $L$  is the usual lag operator used in time-series models—all its elements are equal to zero except those on the first lower diagonal.

The definition of  $\Gamma_L$  is based on the analysis by Kiviet and Phillips (1994) for higher-order time-series models without spatially lagged dependent variables. Note that the  $\Gamma_L$  is not a constant matrix, it is a polynomial in  $L$  and the order of the polynomial depends on the number of time-lagged dependent variables ( $= m_1$ ) and spatially lagged dependent variables ( $= m_2$ ) included in the model.

Next, I use  $u^{**}$  to derive the bias-correction terms in  $\hat{\varphi}_c$ . These terms are given by the expectation of the approximate score function  $S_{NT}^*$ ,

$$S_{NT}^* = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (u_{i,t-1}^{**}, \dots, u_{i,t-m_1}^{**}, W_i u_{i,t-1}^{**}, \dots, W_i u_{i,t-m_2}^{**}, Z_{i,t-1})' (\eta_{it}^d),$$

divided by  $\sqrt{NT}$ . Therefore, I calculate the bias terms using the expectation of

$$[NT]^{-1/2} (u^*)' A \eta, \quad (\text{A.34})$$

where the matrix  $u^*$  is given by

$$u^* = [(L \otimes I_N) u^{**}, \dots, (L^{m_1} \otimes I_N) u^{**}, (L \otimes W) u^{**}, \dots, (L^{m_2} \otimes W) u^{**}, Z]. \quad (\text{A.35})$$

The  $NT \times NT$  matrix  $A$  in (A.34),  $A = [I_T - (i_T)(i_T)' / T] \otimes I_N$ , is an operator that rescales any  $NT \times 1$  vector to have zero mean, i.e.  $A\eta = \eta^d$ . The definition of  $u^*$  in (A.35) follows the work of Kiviet and Phillips (1994).

The (A.34) term does not include the initial conditions of  $u^{**}$ , because the definition of  $u^*$  in (A.35) uses the lag operator  $L$ , which lags any  $T \times 1$  vector and adds a zero as the first element of the lagged vector. For example, the component of  $S_{NT}^*$ , which is related to  $\pi_1$ , includes  $(u_0^{**'}, \dots, u_{T-1}^{**'})$ . The respective component in (A.34) however, includes  $(L \otimes I_N) u^{**} = (\mathbf{0}'_N, u_1^{**'}, \dots, u_{T-1}^{**'})$  ( $\mathbf{0}'_N$  is an  $N \times 1$  vector of zeros).

Next, I calculate the  $(m + K)$  bias correction terms using the expectation of  $[NT]^{-1/2} (u^*)' A\eta$ , where  $u^*$  is given by expression (A.35).

- For the time-lagged dependent variables, the bias terms for the  $\pi_r$  parameters,  $r = \{1, \dots, m_1\}$ , are obtained as follows:

$$\begin{aligned} & [NT]^{-1/2} \mathbf{E} [((L^r \otimes I_N) u^{**})' A\eta] \\ &= [NT]^{-1/2} \mathbf{E} [(u^{**})' (L^r \otimes I_N)' A\eta] \\ &= [NT]^{-1/2} \mathbf{E} [\text{tr} \{ (u^{**})' (L^r \otimes I_N)' A\eta \}] \\ &= [NT]^{-1/2} \mathbf{E} [\text{tr} \{ (L^r \otimes I_N)' A\eta (u^{**})' \}]. \end{aligned}$$

Because  $u^{**} = (\Gamma_L)^{-1} (\tilde{\eta} + u_a^{**})$  and the initial conditions  $u_a^{**}$  are orthogonal to  $\eta$ , the above expression becomes:

$$\begin{aligned} &= [NT]^{-1/2} \mathbf{E} [\text{tr} \{ (L^r \otimes I_N)' A\eta (\Gamma_L^{-1} \tilde{\eta})' \}] \\ &= [NT]^{-1/2} \mathbf{E} [\text{tr} \{ (L^r \otimes I_N)' A\eta (\tilde{\eta})' (\Gamma_L^{-1})' \}] \\ &= [NT]^{-1/2} \mathbf{E} [\text{tr} \{ A\eta (\tilde{\eta})' (\Gamma_L^{-1})' (L^r \otimes I_N)' \}] \\ &= [NT]^{-1/2} \text{tr} \{ A \mathbf{E} [\eta (\tilde{\eta})'] (\Gamma_L^{-1})' (L^r \otimes I_N)' \}. \end{aligned}$$

Because  $\tilde{\eta} = (i_T \otimes c + X\lambda + \eta)$ , the above expression becomes:

$$\begin{aligned} &= [NT]^{-1/2} \text{tr} \{ A \mathbf{E} [\eta (i_T \otimes c + X\lambda + \eta)'] (\Gamma_L^{-1})' (L^r \otimes I_N)' \} \\ &= [NT]^{-1/2} \text{tr} \{ A [(\sigma_\eta^2 + \sigma_{x\eta}\lambda) I] (\Gamma_L^{-1})' (L^r \otimes I_N)' \} \\ &= \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \text{tr} \{ A (\Gamma_L^{-1})' (L^r \otimes I_N)' \} \\ &= \frac{\sigma_\eta^2 + \sigma_{x\eta}\lambda}{\sqrt{NT}} \text{tr} \{ [(L^r \otimes I_N) \Gamma_L^{-1}]' A \}. \end{aligned}$$

In the above expression, the  $1 \times K$  vector  $\sigma_{x\eta} = (\sigma_{1,x\eta}, \dots, \sigma_{K,x\eta})$  contains the expectations  $\sigma_{k,x\eta} = E(X_{k,it}\eta_{it})$ ,  $k = \{1, \dots, K\}$ .

- For the spatially lagged dependent variables, the bias terms for the  $\rho_r$  parameters,  $r = \{1,$



...,  $m_2\}$ , are obtain in a similar way to the bias of the  $\pi_r$  parameters:

$$\begin{aligned}
& [NT]^{-1/2} \mathbf{E} \left[ ((L^r \otimes W) u^{**})' A \eta \right] \\
&= [NT]^{-1/2} \mathbf{E} \left[ \text{tr} \left\{ (u^{**})' (L^r \otimes W)' A \eta \right\} \right] \\
&= [NT]^{-1/2} \mathbf{E} \left[ \text{tr} \left\{ (L^r \otimes W)' A \eta (u^{**})' \right\} \right] \\
&= [NT]^{-1/2} \mathbf{E} \left[ \text{tr} \left\{ (L^r \otimes I_N)' A \eta (\Gamma_L^{-1} \tilde{\eta})' \right\} \right] \\
&= [NT]^{-1/2} \mathbf{E} \left[ \text{tr} \left\{ A \eta (\tilde{\eta})' (\Gamma_L^{-1})' (L^r \otimes W)' \right\} \right] \\
&= [NT]^{-1/2} \text{tr} \left\{ A \mathbf{E} \left[ \eta (i_T \otimes c + X \lambda + \eta)' \right] (\Gamma_L^{-1})' (L^r \otimes W)' \right\} \\
&= [NT]^{1/2} \text{tr} \left\{ A \left[ (\sigma_\eta^2 + \sigma_{x\eta} \lambda) I \right] (\Gamma_L^{-1})' (L^r \otimes W)' \right\} \\
&= \frac{\sigma_\eta^2 + \sigma_{x\eta} \lambda}{\sqrt{NT}} \text{tr} \left\{ [(L^r \otimes W) \Gamma_L^{-1}]' A \right\}.
\end{aligned}$$

- Also, for the endogenous control variables  $X$ , the bias terms for the  $\lambda$  parameters are given by

$$\frac{\sigma'_{z\eta}}{\sqrt{NT}} \text{tr} \{A\},$$

where the  $1 \times K$  vector  $\sigma_{z\eta} = (\sigma_{1,z\eta}, \dots, \sigma_{K,z\eta})$  contains the expectations  $\sigma_{k,z\eta} = \mathbf{E}(Z_{k,it} \eta_{it})$ ,  $k = \{1, \dots, K\}$  ( $Z_k$  is the instrument for the  $k^{th}$  control variable  $X_k$ ).

The bias-corrected estimator for the  $AR(m_1, m_2)$  model is consistent and asymptotically normal under conditions 1 to 4, where the stationarity condition 1(3) becomes

$$\sum_{r=1}^{m_1} |\pi_r| + \sum_{r=1}^{m_2} |\rho_r| < 1.$$

The above condition ensures that the norm of matrix  $\Psi$ ,

$$\Psi = \left[ \sum_{r=1}^{m_1} \pi_r \right] I + \left[ \sum_{r=1}^{m_2} \rho_r \right] W,$$

is bounded and less than one. Also, in the case of the  $AR(m_1, m_2)$ , the maximum column sum of  $W$  should be strictly less than  $k_0$ ,

$$k_0 = \frac{1}{\sum_{r=1}^{m_2} |\rho_r|} \left[ \frac{1}{\sum_{r=1}^{m_1} |\pi_r| + \sum_{r=1}^{m_2} |\rho_r|} - \sum_{r=1}^{m_1} |\pi_r| \right].$$

### A.5 Log-linearization

To derive equation (10), I first log-linearize  $MU_{i,t-1}$  in the left side of (9) as follows:

$$\overline{MU} - \Lambda c_{i,t-1} + \rho_1 \Lambda [W_i c_{t-2}] + \pi_1 \Lambda c_{i,t-2} MU_{i,t-1},$$

where  $\Lambda = \gamma \overline{MU} [1 - \rho_1 - \pi_1]^{-1}$ ,  $\overline{MU} = [1 - \rho_1 - \pi_1]^{-\gamma} (\overline{C})^{-\gamma}$ , and  $\overline{C}$  is the average consumption across U.S. states at the steady state of the economy. Second, I log-linearize  $[\exp(-\beta_i) R_{it} MU_{it}]$  in the right side of (9) as follows:

$$\overline{MU} + \overline{MU} [r_{it} - \beta_i] - \Lambda c_{it} + \rho_1 \Lambda [W_i c_{t-1}] + \pi_1 \Lambda c_{i,t-1}.$$

At the steady state, the product term  $[\exp(-\beta_i) R_{it}]$  is set to 1 to simplify the linearization. Third, I substitute the above two expressions in (9) and rearrange terms to obtain the linearized Euler equation in (10).

**Table 1. Testing Habit Formation Using U.S. State Data for 1966–1998**

The table presents parameter estimates (bold face font) and their t-statistics (below the estimates). Regression 1 uses the LSDV. Regression 7 is estimated with the Anderson and Hsiao (1982) estimator—the instruments for  $\Delta^2 c_t$ ,  $W\Delta^2 c_t$ , and  $\Delta r_t$  are  $\Delta c_{t-2}$ ,  $W\Delta c_{t-2}$ , and  $r_{t-2}$  respectively. Regressions 2 to 6 and 8 to 14 are estimated with the bias-corrected estimator. In regressions 1 to 7 and 13 to 14,  $r_t$  is the only control variable, and its instrument in  $\widehat{\varphi}_c$  is  $r_{t-1}$ . Regressions 8 to 12 include  $r_t$  and an additional variable. Regression 8 includes the growth rates of U.S. services. Regression 9 includes the income growth rate  $\Delta inc_t$ . Regression 10 adds the growth rate of time-lagged income  $\Delta inc_{t-1}$ . Regression 11 includes the squared consumption growth rate. Regression 12 includes the average income growth rate of the states that determine the habit  $W\Delta inc_t$ . Regression 13 is the same as 5 with  $\Delta c$  being replaced by  $\Delta c - \Delta inc$ . The table also reports the within-variation R-squared adjusted for the number of right-hand-side variables in the regression, the curvature parameter  $\gamma$  (its standard error is calculated using the delta method), and the coefficient of relative risk aversion (RRA). The reported RRA is given by the average  $\gamma/S_{it}$ ,  $S_{it} = 1 - (\rho_1 W_i C_{t-1} + \pi_1 C_{i,t-1})/C_{it}$ , across  $t$  and  $i$ .

Type of $W$ :	(1): LSDV Neighbors	(2) Neighbors	(3) Distance	(4) Distance Urban	(5) Distance U > 60%	(6) USA	(7): AH Distance U > 60%
$\Delta c_{t-1}$	<b>-0.02</b> -0.53	<b>0.07</b> 1.60	<b>0.05</b> 1.16	<b>0.04</b> 1.05	<b>0.03</b> 0.66	<b>0.07</b> 1.52	<b>-0.03</b> -0.75
$W\Delta c_{t-1}$	<b>0.29</b> 3.70	<b>0.26</b> 2.50	<b>0.33</b> 2.61	<b>0.34</b> 2.73	<b>0.37</b> 3.14	<b>0.34</b> 2.12	<b>0.60</b> 2.64
$r_t$	<b>0.28</b> 1.82	<b>0.63</b> 3.89	<b>0.61</b> 3.81	<b>0.61</b> 3.77	<b>0.59</b> 3.68	<b>0.62</b> 3.87	<b>0.91</b> 1.50
<i>Curvature</i>	<b>2.60</b> 1.28	<b>1.05</b> 8.27	<b>1.02</b> 7.57	<b>1.02</b> 7.45	<b>1.02</b> 7.31	<b>0.96</b> 6.33	<b>0.48</b> 2.00
<i>RRA</i>	<b>3.55</b>	<b>1.59</b>	<b>1.63</b>	<b>1.65</b>	<b>1.69</b>	<b>1.63</b>	<b>1.10</b>
<i>Adj. R<sup>2</sup></i>	<b>0.06</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.15</b>	<b>0.15</b>
$W = Dist., U > 60\%$	(8) US Serv	(9) $\Delta inc_t$	(10) $\Delta inc_{t-1}$	(11) $(\Delta c_t)^2$	(12) $W\Delta inc_t$	(13) $\Delta c - \Delta inc$	(14) No Neighbors
$\Delta c_{t-1}$	<b>0.02</b> 0.57	<b>0.05</b> 1.15	<b>0.04</b> 0.91	<b>0.03</b> 0.65	<b>0.04</b> 1.01	<b>0.001</b> 0.01	<b>0.05</b> 1.05
$W\Delta c_{t-1}$	<b>0.40</b> 2.58	<b>0.42</b> 2.11	<b>0.40</b> 3.02	<b>0.31</b> 2.17	<b>0.39</b> 2.06	<b>0.34</b> 3.19	<b>0.36</b> 2.86
$r_t$	<b>0.63</b> 2.76	<b>0.66</b> 2.00	<b>0.60</b> 3.62	<b>0.57</b> 3.39	<b>0.62</b> 1.92	<b>0.20</b> 1.55	<b>0.59</b> 3.66
<i>Extra Variable</i>	<b>-0.20</b> -0.26	<b>-0.19</b> -0.28	<b>-0.05</b> -0.50	<b>4.08</b> 0.56	<b>-0.12</b> -0.18		
<i>Curvature</i>	<b>0.92</b> 3.43	<b>0.79</b> 1.82	<b>0.93</b> 5.99	<b>1.16</b> 7.23	<b>0.91</b> 1.67	<b>3.39</b> 0.52	<b>1.00</b> 7.05
<i>RRA</i>	<b>1.59</b>	<b>1.51</b>	<b>1.68</b>	<b>1.77</b>	<b>1.61</b>	<b>5.12</b>	<b>1.68</b>
<i>Adj. R<sup>2</sup></i>	<b>0.17</b>	<b>0.19</b>	<b>0.17</b>	<b>0.31</b>	<b>0.17</b>	<b>0.04</b>	<b>0.16</b>

**Table 2. Simulation Results: No Measurement Error**

The table presents simulation results for estimating  $Y_t = \pi_1 Y_{t-1} + \rho_1 WY_{t-1} + \lambda X_t + c + \eta_t$ . Four cases are simulated for  $T = \{20, 30, 40, 60\}$ . For each case, 20,000 draws are executed. In each draw and for each cross-sectional unit, samples of  $T + 100$  observations are generated. Then, the first 100 observations are deleted. For every draw, the sample of  $T$  observations is used to estimate  $\pi_1$ ,  $\rho_1$  and  $\lambda$ . Three estimators are considered, the LSDV, the bias-corrected  $\hat{\varphi}_c$  (BCORR), and the estimator by Anderson-Hsiao (1982) (AH). Using the 20,000 estimates for  $\pi_1$ ,  $\rho_1$ , and  $\lambda$ , I calculate the percentage mean bias, which is the average of  $[(\hat{\varphi} - \varphi)/\varphi] \times 100$ ,  $\varphi$  and  $\hat{\varphi}$  denote the true and estimated values of  $\pi_1$ ,  $\rho_1$ , and  $\lambda$  (panel A). I also calculate the standard error of  $\hat{\varphi}$  (panel B), and their root mean squared error, which is the square root of the average of  $(\hat{\varphi} - \varphi)^2$  (panel C). For more details on the simulation design see Section 5.1. The values of the standard errors and root mean squared errors in panels B and C are multiplied by a 100.

	<i>LSDV</i>	<i>BCORR</i>	<i>AH</i>	<i>LSDV</i>	<i>BCORR</i>	<i>AH</i>	<i>LSDV</i>	<i>BCORR</i>	<i>AH</i>
<b>Panel A: % Mean Bias</b>									
<i>T</i>	$\pi_1$			$\rho_1$			$\lambda$		
20	-317.7	24.8	10.5	-32.3	-27.9	-22.3	58.5	-16.5	-3.1
30	-264.4	18.1	9.0	-30.4	-27.1	-21.9	54.1	-10.3	-1.7
40	-238.0	15.1	7.6	-29.4	-26.5	-21.7	51.9	-7.6	-1.2
60	-210.9	13.0	6.9	-28.5	-26.0	-21.5	49.7	-5.0	-0.7
<b>Panel B: Standard Error (x100)</b>									
<i>T</i>	$\pi_1$			$\rho_1$			$\lambda$		
20	3.1	3.6	4.8	7.4	7.8	12.6	7.5	11.3	40.1
30	2.5	2.8	3.9	6.0	6.2	10.1	5.7	8.1	28.0
40	2.2	2.4	3.4	5.1	5.3	8.7	4.8	6.6	22.2
60	1.8	1.9	2.7	4.1	4.2	7.0	3.8	5.1	16.5
<b>Panel C: Mean Squared Error (x100)</b>									
<i>T</i>	$\pi_1$			$\rho_1$			$\lambda$		
20	10.0	3.7	4.8	14.1	12.9	15.1	35.3	14.9	40.1
30	8.3	2.9	3.9	12.7	11.8	12.9	32.4	10.1	28.0
40	7.5	2.5	3.4	12.0	11.1	11.8	31.0	8.0	22.2
60	6.6	2.0	2.7	11.3	10.5	10.6	29.6	5.9	16.5

**Table 3. Simulation Results: With Measurement Error**

The table presents simulation results for estimating  $Y_{obs,t} = \pi_1 Y_{obs,t-1} + \rho_1 WY_{obs,t-1} + \lambda X_t^* + c + \eta_t$ .  $Y_{obs,t} = Y_t^* + v_t$ , where  $Y_{obs,t}$  is the observed value of the dependent variable,  $Y_t^*$  is the true value of the dependent variable, and  $v_t$  is the measurement error. There is no measurement error for  $X_t$ ,  $X_t^* = X_{obs,t}$ . Two estimators are considered, the bias-corrected corrected  $\hat{\varphi}_c$  (BCORR), and the estimator by Anderson and Hsiao (1982) (AH). I compute the bias vector, used by  $\hat{\varphi}_c$ , using the AH. In type 1, the instruments for the AH do not suffer from measurement error—they are  $Y_{t-2}^*$ ,  $WY_{t-2}^*$ , and  $X_{t-2}^*$ . In type 2, the instruments for the AH do suffer from measurement error—they are  $Y_{obs,t-2}$ ,  $WY_{obs,t-2}$ , and  $X_{t-2}^*$ . For each type of simulations I calculate the mean bias (panels A1 and A2), standard deviation (panels B1 and B2), and root mean squared error (panels C1 and C2) in the same way as in table 2. The values of the standard errors and root mean squared errors are multiplied by a 100. All other aspects of the simulations are the same as in the case of no measurement error in table 2.

Type 1					Type 2				
AH Instruments: No Measuremt Error					AH Instruments: With Measuremt Error				
	BCORR	AH	BCORR	AH		BCORR	AH	BCORR	AH
Panel A1: % Mean Bias					Panel A2: % Mean Bias				
$T$	$\pi_1$		$\rho_1$		$T$	$\pi_1$		$\rho_1$	
20	-32.6	21.1	-61.7	-20.3	20	-44.2	-60.8	-64.2	-66.9
30	-39.6	16.0	-61.4	-20.6	30	-46.4	-62.4	-63.0	-66.7
40	-43.1	12.7	-61.2	-21.0	40	-47.6	-64.0	-62.4	-66.7
60	-45.1	9.8	-60.9	-21.0	60	-47.8	-65.2	-61.6	-66.5
Panel B1: Standard Error (x100)					Panel B2: Standard Error (x100)				
$T$	$\pi_1$		$\rho_1$		$T$	$\pi_1$		$\rho_1$	
20	7.1	7.5	33.0	20.3	20	3.5	4.7	11.2	16.1
30	2.9	6.0	7.3	16.2	30	2.8	3.8	8.8	13.2
40	2.4	5.2	5.7	13.7	40	2.4	3.3	7.6	11.2
60	1.9	4.2	4.6	11.0	60	1.9	2.7	6.2	9.2
Panel C1: Mean Squared Error (x100)					Panel C2: Mean Squared Error (x100)				
$T$	$\pi_1$		$\rho_1$		$T$	$\pi_1$		$\rho_1$	
20	7.2	7.5	40.2	21.7	20	3.7	5.0	25.1	27.5
30	3.1	6.0	23.9	17.9	30	3.1	4.2	24.2	26.5
40	2.7	5.2	23.4	15.8	40	2.8	3.8	23.8	26.1
60	2.4	4.2	23.0	13.5	60	2.4	3.3	23.3	25.5

## References

- Abel, A. B. 1990, "Asset Prices under Habit Formation, and Catching Up with the Joneses," *American Economic Review Papers and Proceedings*, 80, 38–42.
- Anderson, T. W., and C. Hsiao, 1982, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 45–82.
- Andrews, D., 1991, "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.
- Anselin, L., 2001, "Spatial Econometrics," in B. Baltagi, editor, *A Companion to Theoretical Econometrics*, Oxford: Basil Blackwell, 310–330.
- Asdrubali, P., B. Sorensen, and O. Yosha, 1996, "Channels of Interstate Risk Sharing: United States 1963-1990," *The Quarterly Journal of Economics*, 111, 1082–1110.
- Attanasio, O., and G. Weber, 1995, "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey," *The Journal of Political Economy*, 103, 1121–1157.
- Barro, R. J., and X. Sala-i-Martin, 1992, "Convergence," *Journal of Political Economy*, 100.
- Baltagi, B., S. H. Song, B. C. Jung, and W. Koh, 2004, "Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data," forthcoming in *Journal of Econometrics*.
- Beaudry, P., and E. Van Wincoop, 1996, "The Intertemporal Elasticity of Substitution: An Exploration using a U.S. Panel of State Data," *Economica*, 63, 495–512.
- Boldrin, M., L. J. Christiano, and J. D. M. Fisher, 2001, "Habit Persistence, Asset returns, and the Business Cycle," *American Economic Review*, 91, 149–166.
- Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger, 1989, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance*, 44, 231–262.
- Campbell, J. Y., and J. H. Cochrane, 1999, "By Force of habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–51.
- Campbell, J. Y. and N. G. Mankiw, 1989, "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," in O. J. Blanchard and S. Fischer, editors, *NBER Macroeconomic Annual 1989*, Cambridge, Mass. MIT press, 185–216.
- Carroll, C. D., 1997, "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *Quarterly Journal of Economics*, 112, 1–56.
- , 2000, "Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation)," *Advances in Macroeconomics* 1(1): Article 6.
- Carroll, C. D., J. Overland, D. N. Weil, 2000, "Saving and Growth with Habit Formation," *American Economic Review*, 90, 341–355.

- Case, A., 1991, "Spatial patterns in household demand," *Econometrica*, 59, 953–966.
- Chan, L. Y. and L. Kogan, 2002, "Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *Journal of Political Economy*, 110, 1255–1285.
- Chen, X., and S. C. Ludvigson, 2007, "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," *Working Paper*, New York University.
- Conley, T. G., 1999, "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45.
- Constantinides, G. M., 1990, "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98, 513–43.
- , 2002, "Rational Asset Prices," *Journal of Finance*, 57, 1567–1591.
- Deaton, A., 1991, "Saving and Liquidity Constraints," *Econometrica*, 59, 1221–1248.
- , 1992, "Understanding Consumption," *Clarendon Lectures in Economics*, Clarendon Press, Oxford.
- , 1997, "The Analysis of Household Surveys," *The Johns Hopkins University Press*, Baltimore, Maryland.
- Driscoll, J., and A. Kraay, 1998, "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *The Review of Economics and Statistics*, 80, 549–560.
- Dynan, K., E., 1993, "How prudent are consumers?" *Journal of Political Economy*, 101, 1104–1113.
- , 2000, "Habit Formation in Consumer Preferences: Evidence from Panel Data," *American Economic Review*, 90, 391–406.
- Ferson, W. E., and G. M. Constantinides, 1991, "Habit Formation, and Durability in Aggregate Consumption: Empirical Test," *Journal of Financial Economics*, 29, 199–240.
- Ferson, W., and C. Harvey, 1992, "Seasonality and Consumption-Based Asset Pricing," *The Journal of Finance*, 47, 511–552.
- Flavin, M., 1981, "The Adjustment of Consumption to Changing Expectations About Future Income," *Journal of Political Economy*, 89, 974–1009.
- , 1993, "The Excess Smoothness of Consumption: Identification and Interpretation," *The Review of Economic Studies*, 60, 651–666.
- Fuhrer, J., 2000, "Habit Formation in Consumption and Its Implications for Monetary-Policy Models," *American Economic Review*, 90, 367–390.
- Gali, J., 1994, "Keeping Up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," *Journal of Money, Credit, and Banking*, 26, 1–8.
- Greene, W., 2003, "Econometric Analysis," Prentice Hall, New Jersey.

- Grishchenko, O. A., 2004, "Internal vs External Habit Formation: The Relative Importance for Asset Pricing," *Working Paper*, Pennsylvania State University.
- Grossman, S. J. and R. Shiller, 1981, "The Determinants of the Variability of Stock Market Prices," *The American Economic Review*, 71.
- Hahn, J., and G. Kuersteiner, 2002, "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when both  $n$ , and  $T$  are Large," *Econometrica*, 70, 1639–1657.
- Hahn, J., J. Hausman, and G. Kuersteiner, 2002, "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," *MIT Department of Economics Working Paper*, No. 01-24.
- Hahn, J., and J. Hausman, 2003, "IV Estimation with Valid and Invalid Instruments," *MIT Department of Economics Working Paper*, No. 03-26.
- Hall, R., 1988, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–57.
- Hausman, J., A., 1978, "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.
- Hayashi, F., 1985, "The Permanent Income Hypothesis and Consumption Durability: Analysis Based on Japanese Panel Data," *The Quarterly Journal of Economics*, 100, 1083–1113.
- Heaton, J., 1995, "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," *Econometrica*, 63, 681–717.
- Hernández-Murillo, R., 2003, "Strategic Interaction in Tax Policy Among States," Review, The Federal Bank of St. Louis, May/June Issue, 47–56.
- Hsiao, C., 2003, "Analysis of Panel Data," 2nd Edition, Econometric Society Monographs No. 34, Cambridge University Press.
- Ibragimov, M., 2000, "Conditions for Positive Definiteness of M-Matrices," *Linear, and Multilinear Algebra*, 48, 93–106.
- Islam, N., 1995, "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, 110, 1127–1170.
- Jacobs, K., and K. Wang, 2004, "Idiosyncratic Consumption Risk and the Cross Section of Asset Returns," *The Journal of Finance*, 59, 2211–2252.
- Jagannathan, R., and Y. Wang, 2005, "Lazy Investors, Discretionary Consumption, and the Cross Section of Stock Returns," forthcoming in *Journal of Finance*.
- Kapoor, M., H. Kelejian, and I. Prucha, 2004, "Panel Data Models with Spatially Correlated Error Components," forthcoming in *Journal of Econometrics*.
- Kiviet, J. F., and G. D. A. Phillips, 1994, "Bias Assessment and Reduction in Linear Error-Correction Models," *Journal of Econometrics*, 63, 215–243.
- Kiviet, J. F., 1995, "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models," *Journal of Econometrics*, 68, 53–78.



- Lee, K., H. Pesaran, and R. Smith, 1998, "Growth Empirics: A Panel Approach – A Comment," *Quarterly Journal of Economics*, 113, 319–323.
- Manski, C. F., 1995, "Identification Problems in Social Sciences," Cambridge, Mass.: Harvard University Press.
- Mason, W., and S. Fienberg, 1985, "Cohort Analysis in Social Research: Beyond the Identification Problem," W. Mason and S. Fienberg, editors, New York: Springer Verlag.
- Minc, H., 1988, "Nonnegative Matrices," John Wiley, and Sons.
- Ostergaard, C., B. E. Sorensen, and O. Yosha, 2002, "Consumption, and Aggregate Constraints: Evidence from the U.S. States, and Canadian Provinces," *Journal of Political Economy*, 110, 634–645.
- Phillips, P. C. B. and D. Sul, 2003, "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence," *The Econometrics Journal*, 6, 217–259.
- Potscher, B. M., and I. R. Prucha, 1997, "Dynamic Non-Linear Models, Asymptotic Theory," New York: Springer Verlag.
- Ravina, E., 2005, "Keeping Up with the Joneses: Evidence from Micro Data," *Working Paper*, New York University.
- Shapiro, M. D., 1984, "The Permanent Income Hypothesis and the Real Interest Rate, Some Evidence from Panel Data," *Economics Letter*, 14, 93–100.
- Stock, J. H., J. H. Wright, and M. Yogo, 2002, "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–529.
- Topa, G., 2001, "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies* 68, 261–295.
- Vissing-Jorgensen, A., 2002, "Limited Asset Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.
- Wachter, J., 2006, "A Consumption-Based Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, 76, 362–399.
- Zeldes, S., 1989, "Consumption and Liquidity Constraints: An Empirical Investigation," *The Journal of Political Economy*, 97, 305–346.