The persistent-transitory representation for earnings processes.*

Mette Ejrnæs — Copenhagen University
Martin Browning — Oxford University

November 3, 2013

*We thank an editor and two referees for helpful comments. We also thank the Danish Council for independent research in Social Science (FSE) for financial support.

In the literature on income processes we often seek to decompose shocks into a 'persistent' component which evolves slowly over time and a 'transitory' component which dies away quickly. This scheme is originally due to Kuznets and Friedman (1954) and has been widely used ever since; recent examples include Blundell *et al* (2008), Guvenen (2007), Jappelli and Pistaferri (2010), Meghir and Pistaferri (2004) and Moffitt and Gottschalk (2002). In the panel data literature on dynamic processes the usual approach is to estimate an $ARMA$ model. The permanent-transitory model has number of attractive features compared to the general $ARMA$ model. First, it provides a straight forward interpretation of the shocks. Second, it is possible to assess how the variances of the persistent or transitory variance evolve over time (see e.g. Moffitt and Gottschalk (2002)). Third, the permanent-transitory model make clear how income shock should affect consumption (see Blundell *et al.* (2008)). In the earnings literature, it is almost universally assumed that the persistent and transitory shocks are uncorrelated. This is a very strong assumption. Consider, for example, the loss of a job. This is likely to induce both a persistent effect, if the wage in a new job is lower than in the old job, and a transitory effect, arising from unemployment and a temporary loss of earnings. This in turn gives a positive correlation between the two shocks. This paper lays out the relationship between a generalised version of the permanent-transitory representation and the $ARMA$ approach taking into account the possibility that the persistent and transitory shocks may be correlated.

From the time series literature it is known how the Beveridge-Nelson persistent/transitory decomposition relates to a model similar to the standard

PT model (see Morley *et al* 2003, Oh *et al.* 2008 and Proietti 2006). In the time series literature the permanent-transitory model is known as the *unobserved component decomposition*, in which the permanent part is the *trend* and the transitory component is named the *cyclical innovation*.

From the time series literature it is known that for the permanent-transitory (unit root) model

- Every PT representation has an $ARMA$ representation.
- Every ARMA representation has a PT-representation with correlation of either -1 or +1 between the permanent and transitory shocks.
- The Beveridge-Nelson decomposition also achieves point identification of permanent and transitory shocks by assuming a perfect correlation between the persistent and transitory shocks.

Since evidence is accumulating that no one has an earnings process with a unit root [see, for example, Baker (1997), Guvenen (2009), Browning *et al.*. (2010) and Gustavsson and Osterhold (2010)], we consider a generalization of the usual permanent-transitory model to a persistent-transitory (PT) model. In this model there are two kinds of shocks: a *persistent shock* which has an effect that persists forever (albeit with some possible decay if there is no unit root) and a transitory shock which has only a short run impact (typically one or two periods). For the PT representation, we establish the following:

- The PT representation with uncorrelated shocks implies restrictions on the parameters from an $ARMA$ model.

- If these restrictions are not rejected and it assumed that the persistent and transitory shocks are uncorrelated then the parameters of the PT representation are point identified.
- Without an assumption on the correlation, the parameters of the PT representation are only set identified. The identified set is usually quite wide and, for the leading case, admits *any* correlation between the shocks.
- The ratio of the variances of the persistent and transitory shocks is not point identified if the shocks are correlated, even if we have a unit root.
- Extra information is required to point identify persistent and transitory components. However, even if we can observe the reaction of household consumption to the income shock it requires strong assumptions to actually point identify the parameters.

These results relates to the set identification for which we characterize the set and derive the bounds on important statistics. Also we discuss alternative strategies to point identify the parameters in a PT model.

In the paper we focus on a $ARMA(1,1)$ which is equivalent to assume that the transitory shocks are iid. For this model we can derive analytical results. However, a number of the results can be generalized to models that allow for transitory shocks that are $MA(1)$.[1] In the appendix we show the results for the $ARMA(1,2)$ model. The paper is organized as follows: section 2 outlines the PT representation and section 3 discusses the identification

[1] We acknowledge that the literature also contains more advanced models for the transitory shock, e.g. an $ARMA(1,1)$.

of shocks for the uncorrelated PT representation and the Beveridge-Nelson decomposition. In section 3, we present the general case of a PT model with correlated shocks and derive the identified sets for the correlation and for the ratio of variances. Section 4 discusses how additional information may help to point identify the parameters of the model. Section 5 contains an assessment of the empirical importance of the issues we raise. For a standard data set drawn from the PSID we show that the bounds on the parameters of interest are very wide. In section 6 we present our principal conclusion that the PT representation has to be used with extreme caution.

1 The persistent-transitory representation

Denote log net household income for a given household in period t by y_t. The generalized model[2] is given by the *persistent-transitory* (PT) representation:

$$\begin{aligned} y_t &= \mu + p_t + \tau_t \\ p_t &= \rho p_{t-1} + \eta_t \end{aligned} \tag{1}$$

where p_t is the persistent element and η_t is an *iid*, zero mean shock to the persistent element and τ_t is an *iid*, zero mean transitory shock.[3] We shall assume that the joint distribution of the shocks are characterized by the mean and covariance matrix and denote the (time invariant) variances by σ_η^2 and σ_τ^2, respectively. The covariance between the shocks is denoted $\sigma_{\eta\tau}$. If $\sigma_{\eta\tau} = 0$ we refer to the model as the *uncorrelated PT* representation.[4]

[2] In almost any model of earnings we would want to allow for nonlinear trends. We do not take account of them here to simplify the exposition.

[3] This transitory shock could also include transitory measurement error.

[4] Guvenen (2009) has a similar fomulation of the uncorrelated PT model.

The parameter ρ governs the persistence of the persistent shock; to avoid excessive special cases we shall assume $\rho \in (0, 1]$. That is, we assume that there is some positive persistence. The upper bound, $\rho = 1$, gives the widely used *permanent-transitory* representation:

$$\begin{aligned} y_t &= \mu + p_t + \tau_t \\ p_t &= p_{t-1} + \eta_t \end{aligned} \tag{2}$$

where the parameter μ can, without loss of generality, be suppressed.

2 The identification of persistent and transitory shocks

2.1 The uncorrelated PT representation

To derive the relation between the $ARMA$ model and the persistent-transitory model we consider an $ARMA\,(1, 1)$ model:

$$y_t = \mu\,(1 - \rho) + \rho y_{t-1} + \xi_t + \theta \xi_{t-1} \tag{3}$$

where ξ_t is a zero mean shock with variance ν^2. If $\rho = 1$ this reduces to a unit root with an $MA\,(1)$ stochastic component:

$$\Delta y_t = \xi_t + \theta \xi_{t-1}$$

The PT model (1) always has an $ARMA$ representation which is given by:

$$\begin{aligned} y_t &= \mu + (\rho p_{t-1} + \eta_t) + \tau_t \\ &= \mu + \rho\left(y_{t-1} - \mu - \tau_{t-1}\right) + \eta_t + \tau_t \\ &= \mu\left(1-\rho\right) + \rho y_{t-1} + \eta_t + \tau_t - \rho\tau_{t-1} \end{aligned} \tag{4}$$

Comparing (3) and (4) we see that they have the same deterministic components and only differ in the residual term. We can always point identify ρ and will in the following treat ρ as known.[5] If (3) and (4) are to be equal then:

$$\xi_t + \theta\xi_{t-1} = \tau_t - \rho\tau_{t-1} + \eta_t \tag{5}$$

Most researchers using the PT representation assume that the shocks are uncorrelated ($\sigma_{\eta\tau} = 0$); we make that assumption in this section. Taking the variance and first auto-covariance gives two equations for the mapping between the two sets of parameters[6]:

$$(1+\theta^2)\nu^2 = (1+\rho^2)\sigma_\tau^2 + \sigma_\eta^2 \tag{6}$$

$$\theta\nu^2 = -\rho\sigma_\tau^2 \tag{7}$$

Suppose now that we have estimates of the $ARMA$ parameters (ρ, θ, ν^2) (with $0 < \rho \leq 1$). The following shows the restrictions on these that allow a PT representation.

[5] The parameter ρ is identified from the moment restriction: $E\left(\left(\Delta y_t - \rho\Delta y_{t-1}\right)\left(\Delta y_{t-3} - \rho\Delta y_{t-4}\right)\right) = 0$.

[6] Meghir and Pistaferri (2004) use moments based on a linear combination of variances and first order covariances: $E((\Delta y_{t+1} + \Delta y_t + \Delta y_{t-1}).\Delta y_t)$

Proposition 1 *The ARMA estimates admit an uncorrelated PT representation if and only if:*

$$-\rho \leq \theta \leq 0 \tag{8}$$

Proof. Since $\rho \neq 0$, equations (6) and (7) can be solved to give:

$$\sigma_\eta^2 = (\theta + \rho)(\theta + \frac{1}{\rho})v^2 \tag{9}$$

$$\sigma_\tau^2 = -\left(\frac{\theta}{\rho}\right)\nu^2 \tag{10}$$

To ensure $\sigma_\tau^2 \geq 0$ we require $\theta \leq 0$. For $\sigma_\eta^2 \geq 0$ we have $\theta \geq \max(-\rho, -1/\rho) = -\rho$. ■

Corollary 2 *If the parameter restriction is satisfied then the PT parameters are given by (9) and (10).*

Given the parameter restrictions, we have three important special cases:

1. The permanent-transitory model: The $ARMA$ model with $\rho = 1$, in which case we require $-1 \leq \theta \leq 0$.

2. There are no persistent shocks if $\theta = -\rho$ in which case the $ARMA\,(1,1)$ model is given by:

$$\begin{aligned} y_t &= (1-\rho)\mu + \rho y_{t-1} + \xi_t - \rho\xi_{t-1} \Longrightarrow \\ y_t &= \mu + \rho^t(y_0 - \mu - \xi_0) + \xi_t \end{aligned}$$

3. There are no transitory shocks if $\theta = 0$ in which case the $ARMA\,(1,1)$ reduces to a $AR\,(1)$ model.

Many users of the PT model assume that the transitory component is itself an $MA(1)$ process for which the corresponding $ARMA$ model is an $ARMA(1,2)$ process. The restrictions on the three $ARMA$ parameters that allow us to infer a corresponding PT representation are given in Appendix A.

2.2 The Beveridge-Nelson decomposition.

In time series analysis a Beveridge-Nelson (BN) decomposition is used to decompose a non-stationary process ξ_t into a stationary part and a random walk. In this paper we use the spirit of the BN decomposition to decompose the shock into a transitory and a persistent component; see, for example, Hamilton (1994) section 17.5. For an $ARMA(1,1)$ the 'BN' decomposition is:

$$\begin{aligned} \tau_t &= -\frac{\theta}{\rho}\xi_t \\ \eta_t &= \left(1+\frac{\theta}{\rho}\right)\xi_t \end{aligned} \tag{11}$$

(see the appendix B for the derivation for the more general $ARMA(1,2)$ case). This decomposition has two notable features. First, it does not impose any restrictions on the $ARMA$ parameters since the implied variances will always be non-negative:

$$\begin{aligned} \sigma_\tau^2 &= \left(\frac{\theta}{\rho}\right)^2 \nu^2 \\ \sigma_\eta^2 &= \left(1+\frac{\theta}{\rho}\right)^2 \nu^2 \end{aligned} \tag{12}$$

The other notable feature of this decomposition is that it does *not* impose a zero covariance between the shocks. Instead we have:

$$\sigma_{\eta\tau} = -\frac{\theta}{\rho}\left(1 + \frac{\theta}{\rho}\right)\nu^2 \tag{13}$$

which implies a correlation coefficient of $+1$ if $\theta \in (-\rho, 0)$ and -1 if $\theta > 0$ or $\theta < -\rho$.[7] Thus the identification of the four BN parameters $\left\{\rho, \sigma_\tau^2, \sigma_\eta^2, \sigma_{\eta\tau}\right\}$ from the three $ARMA$ parameters $\{\rho, \theta, \nu^2\}$ is achieved by implicitly assuming a perfect correlation between τ_t and η_t (so long as $\theta \neq 0$ or $\theta \neq -\rho$).

Given that the two procedures for identifying the permanent and transitory shocks from a $ARMA$ process differ in their treatment of the covariance between the shocks, we turn now to the general case.

3 The PT representation with correlated shocks

For earnings processes, the conventional assumption that the transitory and persistent shocks are uncorrelated is very restrictive. Often unemployment is considered to be an important shock to the income process. The change in hours of work is often assumed to be a transitory shock while the changes in hourly wage could be seen as the persistent part of the shock; this clearly induces a positive correlation between the two shocks. Conversely, the correlation would be negative if, for example, a lay-off yields a severance pay which is recorded as a temporary increase in earnings. As another example,

[7] This is immediate from (11) which gives a functional linear relationship between the shocks:

$$\tau_t = \frac{-\theta}{\theta + \rho}\eta_t$$

Hryshko (2013) suggests a negative correlation if being promoted entails losing a bonus. This would be important if the income process is to be used in a consumption simulation model with liquidity constraints.

We will now illustrate how the restrictions on the parameters are affected if we allow for correlated shocks. From equation (5) we have the following two equations that generalise (6) and (7):

$$(1+\theta^2)\nu^2 = (1+\rho^2)\sigma_\tau^2 + \sigma_\eta^2 + 2\sigma_{\eta\tau} \tag{14}$$

$$\theta\nu^2 = -\rho(\sigma_\tau^2 + \sigma_{\eta\tau}) \tag{15}$$

Allowing for correlated shocks removes the restrictions in proposition 1 but this comes at the cost of losing point identification. The following gives the solutions for the correlated PT model:

Proposition 3 *The solutions for the correlated PT model are given by:*

$$\sigma_\tau^2 \in \left[\left(\frac{\theta}{\rho}\right)^2 \nu^2, \left(\frac{1}{\rho}\right)^2 \nu^2\right] \tag{16}$$

$$\sigma_\eta^2 = (1+\theta^2)\nu^2 + \left(\frac{2\theta}{\rho}\right)\nu^2 + \left(1-\rho^2\right)\sigma_\tau^2 \tag{17}$$

$$\sigma_{\eta\tau} = \left(-\frac{\theta}{\rho}\right)\nu^2 - \sigma_\tau^2 \tag{18}$$

Proof. The restrictions on σ_τ^2 are derived from the fact that the variance of σ_η^2 has to be positive and the correlation between η and τ has to lie between -1 and $+1$. ■

Morley *et al.* (2003) and Oh *et al.* (2008) also show the lack of (point) identification by showing how the unobserved components decomposition is related to the BN decomposition. The bounds on the variance of the persistent shock and the covariance are given by:

$$\sigma_\eta^2 \in [\left(\frac{1}{\rho}+\theta\right)^2 \nu^2, \left(1+\frac{\theta}{\rho}\right)^2 \nu^2] \tag{19}$$

$$\sigma_{\eta\tau} \in \left[-\frac{1}{\rho}\left(\frac{1}{\rho}+\theta\right)\nu^2, -\frac{\theta}{\rho}\left(1+\frac{\theta}{\rho}\right)\nu^2\right] \tag{20}$$

Corollary 4 *We can always find a PT representation.*

This follows since the intervals in (16), (19) and (20) are always non-empty if $\theta \in [-1,1]$ and $\rho \in (0,1]$. If $\theta \in (-1,1)$ then point identification fails. For the end points we do have point identification:

Corollary 5 *The PT parameters are all point identified if and only if* $|\theta| = 1$.

The follows since the interval in (16) is a point if and only if $\theta^2 = 1$. The final corollary is that the ratio of variances is generally not point identified.

Corollary 6 *If* $\theta \neq 0$*, the bounds on the ratio of persistent variance to the transitory variance are given by:*

$$\frac{\sigma_\eta^2}{\sigma_\tau^2} \in [(1+\theta\rho)^2, \left(1+\frac{\rho}{\theta}\right)^2] \tag{21}$$

This follows since the ratio of the variances is given by

$$\frac{\sigma_\eta^2}{\sigma_\tau^2} = \frac{(1+\theta^2)\nu^2}{\sigma_\tau^2} + \frac{2\theta\nu^2}{\rho\sigma_\tau^2} + (1-\rho^2). \tag{22}$$

The ratio is an decreasing function of σ_τ^2 and the bounds follow from the bounds on σ_τ^2.

One implication of (21) is that the ratio if the correlation is set to zero (given by the ratio of the variances in (9) and (10)) is the harmonic mean of the end points of the identified set given in (22).

The fact that the ratio between the persistent variance to the transitory variance is not point identified has also been discussed intensively in the literature on decomposition of GDP (see Morley *et al* (2003)). Here they find that the BN decomposition and a uncorrelated PT model give very different results in terms of characterizing the cyclical innovations and the trends.

For any solution, the correlation between the shocks is given by:

$$\chi_{\eta\tau} = \frac{\sigma_{\eta\tau}}{\sqrt{\sigma_\eta^2}\sqrt{\sigma_\tau^2}}$$

This is not point identified; the following gives the bounds on the identified set.

Proposition 7 *If $-\rho < \theta < 0$ then we can find a solution to (14) and (15) with $\chi_{\eta\tau} \in [-1, 1]$.*

If $\theta > 0$ or $\theta < -\rho$ then we have $\chi_{\eta\tau} \in [-1, \lambda]$ where

$$\lambda = -2\frac{\sqrt{\frac{\theta}{\rho}(\theta + \rho)(\theta + 1/\rho)}}{|1 + \theta^2 + (2\theta)/\rho|}. \tag{23}$$

Proof. To show this we assume, without loss of generality, that $\nu^2 = 1$. Suppose $-\rho < \theta < 0$. Taking the lower bound in (16) we have:

$$\begin{aligned}
\sigma_\tau^2 &= \left(\frac{\theta}{\rho}\right)^2 \\
\sigma_\eta^2 &= (1 + \theta^2) + \left(\frac{2\theta}{\rho}\right) + \left(1 - \rho^2\right)\left(\frac{\theta}{\rho}\right)^2 = \left(1 + \frac{\theta}{\rho}\right)^2 \\
\sigma_{\eta\tau} &= -\frac{\theta}{\rho} - \left(\frac{\theta}{\rho}\right)^2 = -\frac{\theta}{\rho}\left(1 + \frac{\theta}{\rho}\right) \\
\chi_{\eta\tau} &= \frac{-\frac{\theta}{\rho}\left(1 + \frac{\theta}{\rho}\right)}{\left(1 + \frac{\theta}{\rho}\right)\left(-\frac{\theta}{\rho}\right)} = 1
\end{aligned}$$

Note that we take $\sqrt{\sigma_\tau^2} = -\theta/\rho$ to ensure the standard deviation is non-negative. If we take the upper bound in (16) we can show that $\chi_{\eta\tau} = -1$. The correlation coefficient is a continuous function of σ_τ^2 between these bounds and hence any value of σ_τ^2 corresponding to an arbitrary $\chi_{\eta\tau} \in [-1, 1]$ is a solution. To prove the second statement, both the lower and upper bound for $\sigma_{\eta\tau}$ in (20) always give a correlation of -1. We can find the upper bound for the correlation in the case where $\theta > 0$ or $\theta < -\rho$. The correlation is given by

$$\chi_{\eta\tau} = \frac{-(\theta/\rho) - \sigma_\tau^2}{\sqrt{\left(1+\theta^2\right) + (2\theta)/\rho + (1-\rho^2)\sigma_\tau^2}\sqrt{\sigma_\tau^2}}.$$

We can find the maximum correlation by solving the first order condition. The solution to the first order condition is

$$\sigma_\tau^2 = \frac{\theta}{\rho}\frac{\left(1+\theta^2+(2\theta)/\rho\right)}{1+\theta^2+2\theta\rho}$$

The maximum correlation is

$$\lambda = \chi_{\eta\tau} = -2\frac{\sqrt{\frac{\theta}{\rho}(\theta+\rho)(\theta+1/\rho)}}{|1+\theta^2+(2\theta)/\rho|}$$

■

These are striking results. The first statement implies that if $\theta \in [-\rho, 0]$ there are no bounds on the correlation between shocks.[8] The case with no correlation corresponds to the usual PT identifying assumption whereas the case with a correlation of $+1$ corresponds to the BN decomposition. The various cases are illustrated in the top panel of figure 1. The second

[8] The coincidence of this with the parameter values that yield the uncorrelated PT representation (proposition 1) is particular to the $ARMA(1,1)$ case. For example, it does *not* hold for the $ARMA(1,2)$ case.

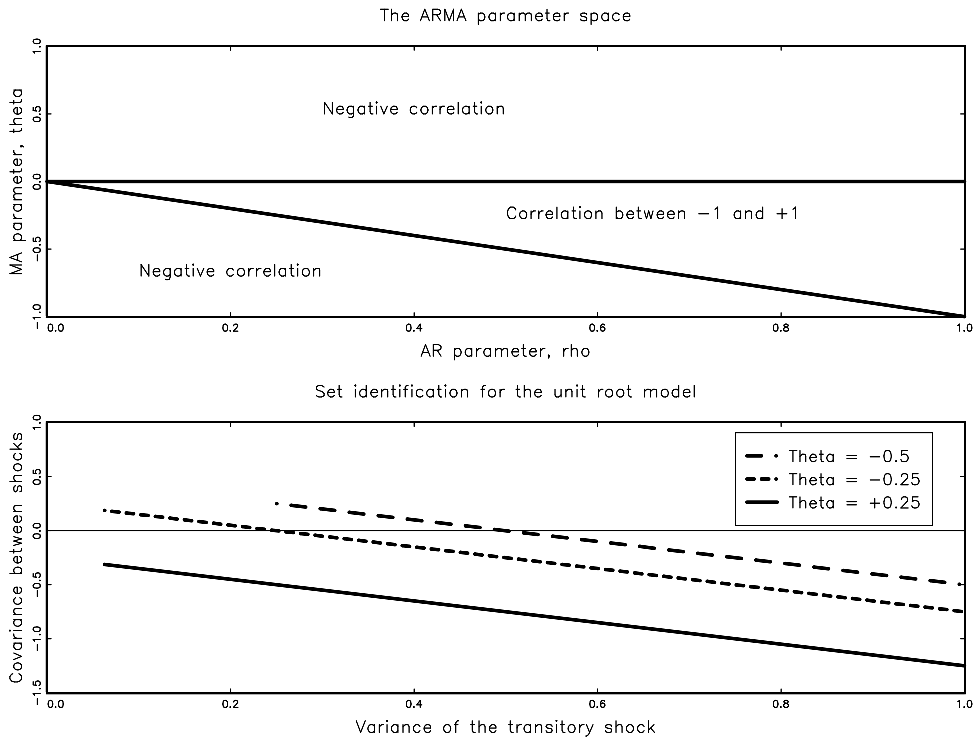


Figure 1: Set identification for the ARMA model

statement also has significant implications: if θ is outside the bounds given for the uncorrelated PT representation then the covariance between the two shocks is always negative, with perfect negative correlation (the BN case) always being a solution. However, having a negative correlation is exactly the case that we would often wish to exclude. The bounds get tighter as θ approaches -1 or $+1$.[9] In figure 2, we show the upper bound for four different values of ρ.

[9]This also follows from corollary 5 which states that the parameters are point identified if $|\theta| = 1$.

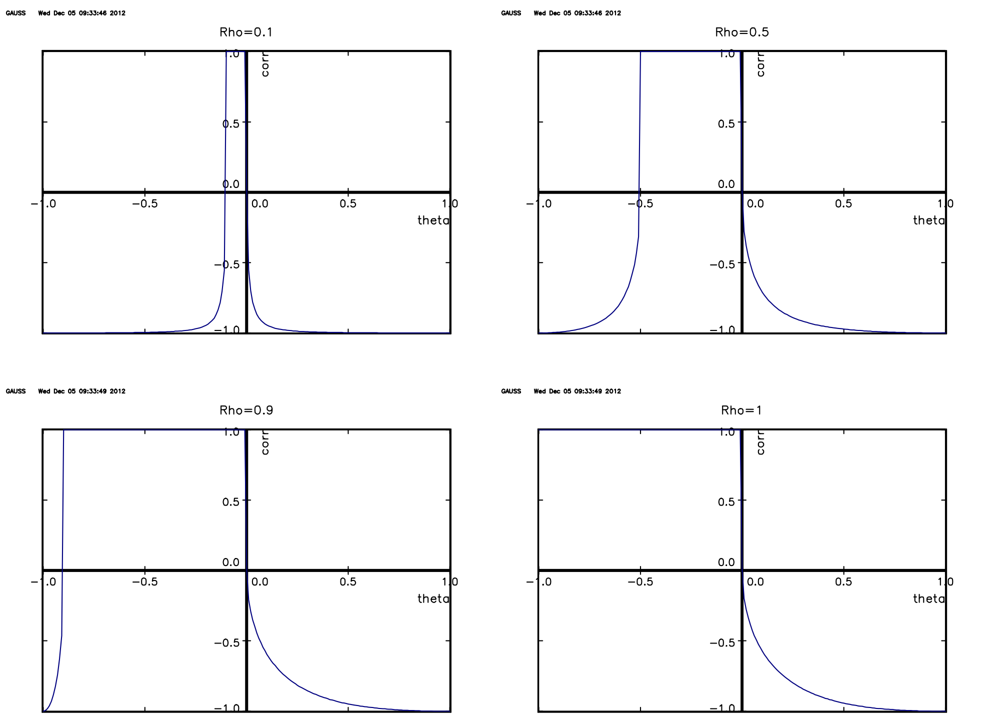


Figure 2: Bounds on the correlation for different values of ρ

Although not all of the parameters of the PT representation are point identified for the unit root model ($\rho = 1$), the variance of the permanent shock is point identified with $\sigma_\eta^2 = (1+\theta)^2 \nu^2$; see equation (17). However, the two remaining parameters σ_τ^2 and $\sigma_{\eta\tau}$ are only set identified. In the bottom panel of figure 1 we show the set identification for a unit root model ($\rho = 1$) with different values of θ and $\nu^2 = 1$. In many applications of the unit root permanent-transitory model, the ratio of the variance of the transitory shocks to the variance of the permanent shocks is of importance. These calculations show that this ratio is not identified and can vary substantially; for example, if $\theta = -0.5$, the ratio can be between unity and four.[10]

To illustrate the importance of the non-identification we have constructed two series of transitory and permanent shocks ($\rho = 1$), which generate exactly the same y process. Both sets of transitory and permanent shocks satisfy the conditions stated in section 1.[11] The first set is with uncorrelated transitory and permanent shocks; this is shown in the upper panel of the figure 3. The lower panel of the figure 3 corresponds to transitory and permanent shocks that are perfectly positively correlated (the BN decomposition). The figure shows that two very different set of shocks can generate with exactly the same time series. Note that the variance of the transitory shocks is much lower when the shocks are correlated. This also shows that the non-identification cannot be resolved by including additional moments of the y-process. If we want to obtain point identification we need additional assumptions on the process and in some cases also additional moments.

[10] This follows from corollary 6.

[11] A description of how the figures are generated can be found in appendix C.

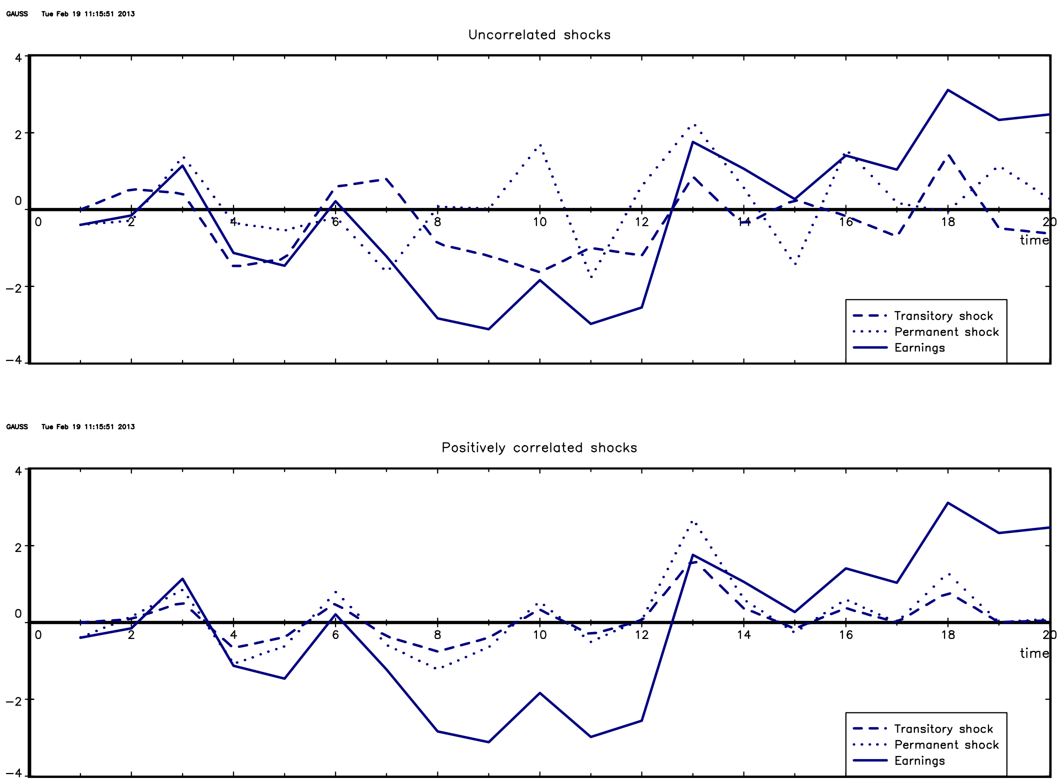


Figure 3: PT model with uncorrelated shock and positively correlated shock

4 Alternative identification strategies

In the previous section we have shown that in the general PT model with correlated shocks the variances are only set identified. Point identification requires either stronger assumptions or additional information or data. In this section we show how either additional structure imposed by the theory, consumption data or applying higher order moments can help in point identifying the parameters. However, all the alternative strategies require strong additional assumptions.

4.1 Identification using structural wage and labour supply models

One alternative way of identifying the variances is by imposing more structure on the earnings process by using theory on labour supply, wages and earnings. By using a (semi) structural model on labour supply, wages and earnings one can decompose the shocks into transitory and persistent shocks. Altonji *et al.* (2013) explicitly model both labour supply and the accumulation of tenure together with wages and earnings. In such a model, the event of losing a job will generate both a transitory and a persistent effect on earnings. The transitory effect arises because of the reduction in labour supply while the loss of tenure generates a persistent effect on the wage and thereby affects earnings. By using the structure imposed by the wage and labour supply model one may be able to identify an observable event such as displacement, that can be seen as a transitory and/or persistent shock and pin down the correlation between the two shocks. Other potentially observable events include promotion, a temporary lay-off, a health shock, occupational mobility and geographical mobility.

4.2 Information on expectations

An alternative route to point identification is given by Pistaferri (2001) who suggests using additional information in a panel of income and wealth (the Survey of Italian Households' Income and Wealth (SHIW)). The extra information is a direct question, asked in period t, about expected income growth between t and $t+1$. This is interpreted as giving a direct measure of $E_t(\Delta y_{i,t+1})$, the conditional mean of income growth conditional on information at time t. Applying this to the permanent-transitory model (2)[12] we have:

$$\begin{aligned} E_t(\Delta y_{i,t+1}) &= E_t(p_{t+1} + \tau_{t+1} - p_t - \tau_t) \\ &= E_t(\eta_{t+1} + \tau_{t+1} - \tau_t) = -\tau_t \end{aligned} \tag{24}$$

Thus the conditional mean of income growth yields an estimate of the (negative) transitory shock itself. Given this, the permanent shock can be recovered from:

$$\begin{aligned} \Delta y_{it} + E_t(\Delta y_{i,t+1}) - E_{t-1}(\Delta y_{it}) &= (\eta_t + \tau_t - \tau_{t-1}) - \tau_t + \tau_{t-1} \\ &= \eta_t \end{aligned} \tag{25}$$

Two points arise. First, the informational requirement here is much higher than in most income panels. Even in the SHIW, complications arise

[12] Similar manipulations can be carried out for the persistent-transitory model but for clarity we here take the model with $\rho = 1$.

since the survey is run every second year but the expectations question only asks about the next year. Having only $(y_{i,t+2} - y_{it})$ requires supplemental assumptions to identify τ_t directly. More importantly, the method requires that the expectations measure be free of measurement error. It can be shown that measurement error in the expectations response will lead to a downward bias in the estimate of the correlation between the two shocks.

Pistaferri (2001) uses these in a model of consumption and income but does not present an estimate of the correlation between the shocks. Hryshko (2013) adopts the Pistaferri method and using the Italian SHIW calculates the correlation between the persistent shock and the permanent shock. He finds a strong negative correlation.

4.3 Identification using consumption data

If consumption data are available this can be used to point identify the parameters. We assume a setup as in Blundell, Pistaferri and Preston (2008) in which consumption changes depend on the transitory and persistent shock. To simplify the exposition we assume $\rho = 1$ with a model for income and consumption growth given by:

$$\begin{aligned} \Delta y_t &= \eta_t + \tau_t - \tau_{t-1} \\ \Delta c_t &= \phi\eta_t + \psi\tau_t + \upsilon_t. \end{aligned} \tag{26}$$

Using the moment conditions given in the appendix of Blundell *et al* (2008) we have:

$$\begin{aligned} E((\Delta y_{it+1} + \Delta y_{it} + \Delta y_{it-1})\Delta y_{it}) &= \sigma_\eta^2 \\ E(\Delta y_{it}\Delta y_{it-1}) &= -\sigma_\tau^2 - \sigma_{\eta\tau} \\ E((\Delta y_{it+1} + \Delta y_{it} + \Delta y_{it-1})\Delta c_{it}) &= \phi\sigma_\eta^2 + \psi\sigma_{\tau\eta} \\ E(\Delta c_{it}\Delta y_{it+1}) &= -\psi\sigma_\tau^2 - \phi\sigma_{\tau\eta} \\ V(\Delta c_{it}) &= \phi^2\sigma_\eta^2 + \psi^2\sigma_\tau^2 + 2\phi\psi\sigma_{\tau\eta} + V(\upsilon_t). \end{aligned} \quad (27)$$

The equations show that the variance of the persistent shock, σ_η^2, is point identified (as it is for any unit root model). Using the second to fourth moments we cannot point identify $(\sigma_\tau^2, \sigma_{\eta\tau}, \phi, \psi)$.[13] To point identify the remaining parameters we need at least one of the following conditions to be imposed:

$$\begin{aligned} \sigma_{\tau\eta} &= 0 \\ \psi &= 0 \\ \phi &= 0 \end{aligned}$$

The first restriction is uncorrelated shocks (as assumed in Blundell *et al* (2008)). The second restriction states that there is no impact of transitory shocks on consumption; this is the most natural assumption. The third restriction is that the counter-intuitive assumption that there is no impact of permanent shocks on consumption.

[13] Where the consumption information is directly useful for estimating income processes is in identifying measurement error in income. In a model with only income information, classical measurement error is indistinguishable from the transitory shock.

Hryshko (2013) shows what happens in a consumption model such as (26) if we mistakenly assume zero correlation (as in most studies of income and consumption). In the case where there is a negative correlation ($\sigma_{\tau\eta} < 0$) this will lead to a downward bias in ϕ and ψ. In this case one will find 'excess smoothness of consumption'. On the other hand, if the correlation is positive the estimates of ϕ and ψ are upward biased. Furthermore, Hryshko (2013) argues that a consumption-income model which allows for negatively correlated permanent and transitory shocks in the income process better explains the 'excess smoothness' which is often found in consumption data. Hryshko (2013) use a structural model of income and consumption (with assumption imposed by the structural relation) and is thereby able to point identify the parameters. He estimates the correlation between the transitory and the permanent income shock to be -0.6.

4.4 Identification from higher order moments

Finally, we show how assumptions on higher order moments can help to pin down the correlation between the transitory and persistent shock. Assume that we have a correlated PT model. We can write the persistent and transitory shocks as:

$$\begin{aligned} \eta_t &= z_t + v_t \\ \tau_t &= a z_t + u_t. \end{aligned}$$

where z_t is the common part of the shock. The parameters of the shocks are given by a, σ_z^2, σ_u^2 and σ_τ^2, where a identifies the covariance. For known ρ we have:

$$\begin{aligned} y_t - \rho y_{t-1} &= \eta_t + \tau_t - \rho\tau_{t-1} \\ &= (1+a)z_t + v_t + u_t - \rho a z_{t-1} - \rho u_{t-1}. \end{aligned}$$

If z_t is drawn from an asymmetric distribution but u and v are symmetric we can identify the a (and thereby the covariance). Assuming:

$$\begin{aligned} E(z_t^3) &\neq 0 \\ E(u_t^3) &= E(v_t^3) = 0. \end{aligned}$$

we have:

$$E((y_t - \rho y_{t-1})^3) = ((1+a)^3 - \rho^3 a^3)E(z_t^3),$$

and we can then identify a if $E(z^3)$ is a known or a function of first and second order moments. Similarly we can identify the covariance if either $E(u_t^3)$ or $E(v_t^3)$ was different from zero. Higher order moments such as fourth order moments can also be used but again it requires that $E(z^4)$ is known or at least a known function of first and second order moments.

5 Quantitative implications

5.1 Empirical specification

The analysis above shows that for given values of the $ARMA$ parameters, the bounds on the estimates of the ratio of the persistent and transitory variances can be quite wide. In this section we consider whether the possibility of wide bounds is actually realised for a given sample of workers. To quantify the implications, we follow closely Browning, Ejrnæs and Alvarez (2010) (BEA); readers are referred to that paper for a detailed rationale of the empirical approach we use here. BEA estimate an $ARMA(1,2)$ model with a quadratic trend and an allowance that the reversion to the trend (the auto-regressive parameter) is time dependent. To make our point cleanly, we here take a simplified version of this model with an $ARMA(1,1)$ process with a linear trend and a time independent AR parameter. For household i $(= 1, ...H)$ in time t $(= 2, ..T)$, log earnings, y_{it}, are given by:

$$y_{it} = [\mu_i(1-\rho_i) + \rho_i\alpha_i] + \rho_i y_{i,t-1} + [\alpha_i(1-\rho_i)](t-1) + \nu_i\xi_{it} + \theta_i\nu_i\xi_{i,t-1} \quad (28)$$

where the ξ_{it}'s are independent standard normals. If $\rho_i = 1$ this reduces to a unit root model with a drift equal to α_i. The important element of this specification is that we allow that there is pervasive heterogeneity; that is, all of the parameters $(\mu, \alpha, \rho, \theta, \nu)$ are allowed to vary across workers. Apart from the gain in the fit to the data, this is particularly useful in the current context since it allows us to examine the cross-section distribution of the identified sets we are interested in.

To start the process, we model the initial condition by:

$$y_{i1} = a_0 + a_1 d_i + c_0\xi_{i1} + c_1\xi_{i0} \quad (29)$$

where d_i is the year of birth of worker i to allow for age/cohort effects in the starting value; again, see BEA.

To model the heterogeneity we adopt a two factor structure for the parameters in (28).[14] Letting η_{ki} be independent standard normals for $k = 1, 2$

[14] This specification is the end result of a specification search that began with a five

and $i = 1, ..H$, we take:

$$\begin{aligned} \mu_i &= \phi_1 + \exp(\psi_{11}) \eta_{1i} \\ \nu_i &= \exp(\phi_2 + \psi_{21}\eta_{1i} + \exp(\psi_{22}) \eta_{2i}) \\ \alpha_i &= \phi_3 + \psi_{31}\eta_{1i} + \psi_{32}\eta_{2i} \\ \rho_i &= \ell(\phi_4 + \psi_{41}\eta_{1i} + \psi_{42}\eta_{2i}) \\ \theta_i &= 2\ell(\phi_5 + \psi_{51}\eta_{1i} + \psi_{52}\eta_{2i}) - 1 \end{aligned} \tag{30}$$

where $\ell(x) = \exp(x) / (1 + \exp(x))$ so that $\rho \in (0, 1)$ and $\theta \in (-1, 1)$.[15] This structure allows for a good deal of heterogeneity with dependence across parameters. In all we have 18 parameters to estimate: (a_0, a_1, c_0, c_1) from (29) and $(\phi_1, ..\phi_5, \psi_{11}, ...\psi_{52})$.

5.2 Estimation method

We use indirect inference to estimate the parameters governing the distribution of parameters. Gouriéroux, Phillips and Yu (2010) provide a strong defence for using indirect inference to estimate the parameters of a parametric dynamic panel model. The main motivation is that this method provides a bias reduction estimation method to allow for the well known bias in dynamic panel data estimation.

Indirect inference requires us to specify auxiliary parameters (ap's) that can be calculated on the data to hand and on simulated data that purports to

factor model. The $\chi^2(6)$ test statistic for dropping three factors was 3.6.

[15] The restriction on ρ explicitly rules out that anyone has a unit root. Guvenen (2009) and Browning et al. (2010) provide evidence that this is not rejected if we allow for heterogeneous deterministic trends (the α_i's).

model the empirical generating process. Indirect inference chooses parameter estimates to minimise the distance between the two sets of auxiliary parameters. In our estimation procedure we rely heavily on auxiliary parameters that are based on regressions for each worker; this follows the literature on testing for unit roots in panel data (see Levin *et al* (2002) or Im *et al* (2003)). Of course, the estimates from individual regressions based on short time series do not give unbiased estimates of anything of interest. However, the use of the same auxiliary process for the data sample and the simulated sample introduces similar biases in both; it is in this sense that Gouriéroux, Phillips and Yu (2010) see indirect inference as a bias reduction method. As well as auxiliary parameters based on individual regressions we also use moments that have been widely used in the earnings process literature. Details of the construction of the auxiliary parameters are given in appendix D. These statistics provide a rich description of the time-series and cross-section features of the original data. In all we have 42 auxiliary parameters for the 24 distribution parameters to be estimated.

To estimate we have to simulate from the model in (28) and (29). To reduce the impact of the misspecification in the initial value we start the process at $t = -3$ (using (29)) and then recursively generate $y_{i,t-2.}, ..y_{iT}$. We then discard the first three observations for each worker. Additionally we have to allow that the panel we use is unbalanced with some workers not in the first observation period and some dropping out before the final observation period. To take account of this, we replicate the actual data with the values not observed for each replicated household masked out as in the data. For example, if household i is only observed for periods 4 to 16 then for each replicated household for that particular household we simulate from 1 to T and set the periods $1-3$ and $17-T$ as missing. Thus the simulated data has the same imbalance as the original data.

5.3 Sample

We estimate using a subsample of the sample drawn from the PSID as used in Meghir and Pistaferri (2004) (MP). This is an unbalanced sample of male workers followed from (survey years) 1968 to 1993. We select on being aged between 25 to 57 and being in the sample for at least nine years. The original MP sample consists of $2,069$ individuals, with $31,631$ observations. The earnings variable includes all income from labour, deflated to the year 1992. For individuals in this sample the variables we use for controlling for observable heterogeneity are education, race, age and birth cohort. We deal with some of the observable heterogeneity by stratifying on education and working by selecting whites with a high school education. This gives a sample size of 749 with workers being observed between 9 and 26 years, which gives in total $11,503$ observations. Following MP, we run a first round regression of log earnings on year dummies and age dummies and treat the residuals from this regression as earnings in (28).

5.4 Empirical estimates

The $\chi^2(24)$ test statistic for the over-identifying restrictions has a value of 38.3 with an associated p-value of 2.9%. Although marginal, the fit is quite good and we deem it unlikely that a marginal improvement in the fit from generalising the model would change the qualitative implications below. The

Parameter	10%	25%	50%	75%	90%
μ	-0.31	-0.20	-0.09	0.02	0.13
α	-0.10	-0.07	-0.03	0.00	0.04
ρ	0.48	0.66	0.82	0.91	0.96
θ	-0.48	-0.37	-0.22	-0.07	0.07
ν	0.08	0.11	0.16	0.22	0.31

Table 1: Marginal distributions of model parameters

$\chi^2(4)$ test statistic for shutting down the heterogeneity in (ρ,θ) (that is, imposing $\psi_{41} = \psi_{42} = \psi_{51} = \psi_{52}$) is 24.0 which indicates that there is significant heterogeneity in the $ARMA$ parameters. Table 1 presents the marginal distributions of the heterogeneous model parameters.[16] As can be seen, all of the parameters are quite dispersed. Since these distributions are similar to those in BEA, we only discuss the $ARMA$ parameters that are the focus of this study. The AR parameter is quite dispersed with no significant bunching near unity (see BEA for details of how this relates to a test for anyone having a unit root). The MA parameters are mostly negative, but a considerable proportion have a positive value.

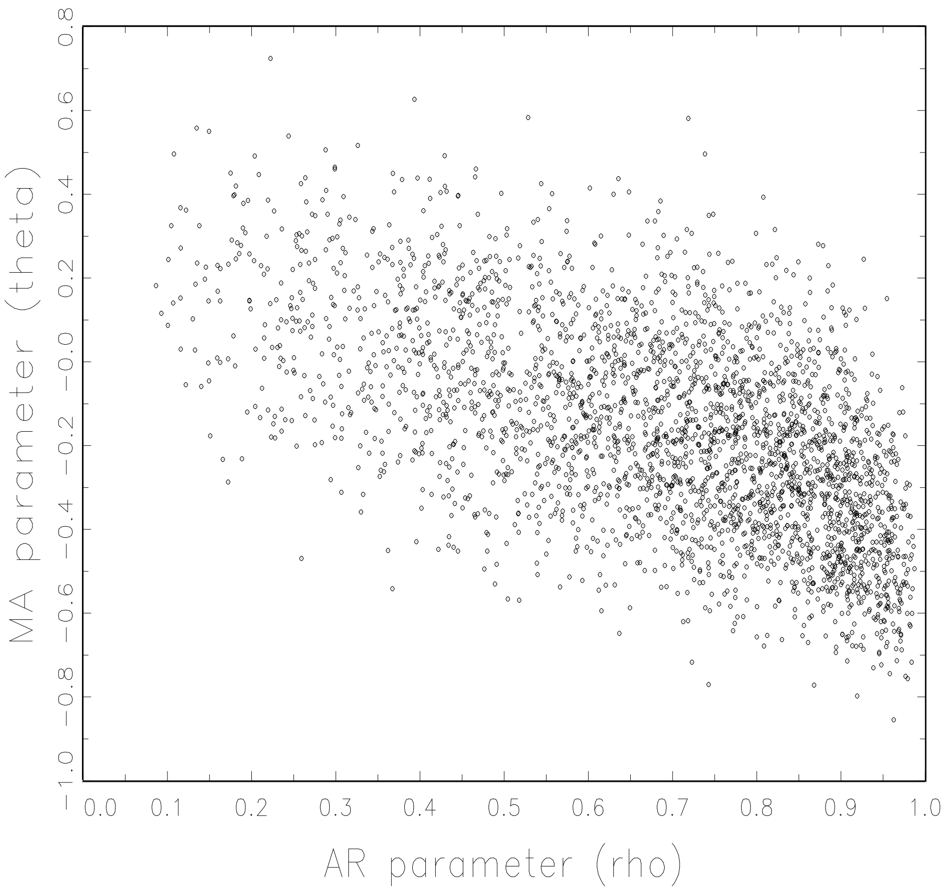
Table 2 presents the correlations of the heterogeneous model parameters. This shows that there is considerable dependence between the model parameters; this is to be expected in a low dimensional factor model. Of particular interest for us, the $ARMA$ parameters ρ and θ are strongly negatively correlated; this is also clear from the scatter plot of the estimates in figure 5.4. Most of the estimates have $-\rho < \theta < 0$ so that for a majority of the pop-

[16] Estimates of the parameters in (29) and (30) are given in appendix D.2.

	μ	α	ρ	θ	ν
μ	1	–	–	–	–
α	-0.95	1	–	–	–
ρ	-0.6	0.39	1	–	–
θ	0.26	-0.15	-0.57	1	–
ν	0.93	-0.88	-0.58	0.24	1

Table 2: Correlations of model parameters

ulation any correlation between -1 and $+1$ is admissible. However, for a significant proportion (mostly with $\theta > 0$) the PT representation only allows a negative correlation or no representation if we impose zero correlation.



The joint distribution of the ARMA parameters

5.5 Implications

Finally we can turn to the implications of these estimates for the identification issues we are primarily concerned with. We first present the implications for the mean of the $ARMA$ parameters which are $\rho = 0.69$ and $\theta = -0.17$. These estimates are unremarkable for (stationary) models that assume no heterogeneity in the $ARMA$ coefficients. The restrictions from proposition 1 require $-\rho \leq \theta \leq 0$ so that at the mean, the $ARMA$ model admits a

PT representation with zero correlation between the shocks. It thus follows that the identified set for the correlation between the persistent shock and the transitory shock is $[-1, 1]$. With these mean values, the identified set for the ratio of the variance of the persistent shock to the variance of the transitory shock is $[0.70, 6.96]$ so that the highest possible ratio is 10 times the lowest possible. If we set the correlation to zero, we have a point estimate of 2.21 (the harmonic mean of the end points of the identified set). Comparing to the literature which all assume zero correlation, Moffitt and Gottshalk (1994) and Guvenen (2009) also find the ratio of the variances close to 2, while Meghir and Pistaferri (2004), Blundell *et al.* (2008) and Hryshko (2013) find the ratio in the range $0.5 - 1$.[17]

Having allowed for heterogeneous $ARMA$ parameters we can also display the distributions of the identified sets. For all values of (ρ, θ) we can construct the ratio of the variance of the persistent shock to the variance of the transitory shock from (21). The first row of Table 3 shows the distribution of the ratio of the upper bound to the lower bound for the ratio of variances. The median value is 8.5 which is slightly lower than the ratio given above at the mean of the estimates but still very high. Since the upper bound is unbounded above ($\sigma_\eta^2/\sigma_\tau^2 \to \infty$ as $\theta \to 0$) we see very high values for the upper tail. The second row of Table 3 shows the upper bound on the correlation for the 17.7% of the sample who do not satisfy the condition that $-\rho \leq \theta \leq 0$ (as shown above the lower bound is always -1). Even in the tail of this distribution, the identified set is quite wide.

[17] The ratio when we assume zero correlation is very sensitive to the value taken for ρ.

	10%	25%	50%	75%	90%
$\sigma_\eta^2/\sigma_\tau^2$	1.2	3.1	8.5	38.0	219.6
Upper bound for correlation	-0.91	-0.82	-0.65	-0.48	-0.34

Table 3: The distribution of identifed sets

6 Conclusion

This paper has derived the relationship between $ARMA$ estimates of a dynamic model and the representation that allows us to identify a transitory and a persistent component for shocks. The main result is that the decomposition is critically dependent on the assumed correlation between the two shocks. In the absence of such an assumption, the parameters of the PT representation are only set identified. Moreover, there are no bounds on the correlation between the shocks if we have a moderate negative MA parameter. A quantitative assessment of the seriousness of this lack of point identification suggests that it is very serious with very wide bounds for both the correlation and the ratio of the variances of the shocks.

The non-identification has important implications for analyses made on earnings dynamics. For example, analyses that examine how the variance of transitory and persistent shocks have changed over time are only valid under the assumption that the correlation remains unchanged. Thus a finding on a time varying ratio of variances (see Moffitt Gottschalk (2012)) could be generated by constant variances and time varying covariances. An apparent increase in the transitory variance in a uncorrelated PT model could instead have been induced by have been decrease in the correlation. Also analyses that examine the ability of households to smooth consumption are heavily influenced by the assumption on the correlation with wildly varying values for the variance of the persistent shock relative to the transitory shock. Given this, it will be advisable for future researchers to present identified sets for outcomes of interest that explicitly take account of the non-identification of the correlation between the shocks.

References

[1] Altonji, Joseph G.; Anthony Smith and Ivan Vidangos. 2013. "Modeling Earnings Dynamics", *Econometrica* 81(4), pp. 1395-1454.

[2] Baker, M. (1997). "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings", *Journal of Labor Economics*, 15(2), pp. 338-75.

[3] Blundell, R., L. Pistaferri and I. Preston (2008), "Consumption Inequality and Partial Insurance." *American Economic Review* 98(5): 1887-1921.

[4] Browning, M.; M. Ejrnæs and J. Alvarez. 2010. "Modelling Income Processes with Lots of Heterogeneity", *Review of Economic Studies*, 77(4), pp. 1353-81.

[5] Friedman, Milton and Simon Kuznets (1954), *Income from Independent Professional Practice*, National Bureau of Economics, NBER.

[6] Gouriéroux, C.; P. C. B. Phillips and J. Yu. (2010). "Indirect Inference for Dynamic Panel Models", *Journal of Econometrics*, 157(1), pp. 68-77.

[7] Gustavsson,M and P. Osterholm (2010), "*Does Labor-Income Process have a Unit Root? Evidence from Individual-Specific Time Series"*. Working paper 2010:21, Uppsala University

[8] Guvenen, F. (2007), "Learning your earning: Are labor income shocks really very persistent?" *American Economic Review* 97(3): 687-712.

[9] Guvenen, F. (2009). "An Empirical Investigation of Labor Income Processes", *Review of Economic Dynamics*, 12(1), pp. 58-79.

[10] Hamilton, James D (1994) *Time series analysis.* Princeton University Press.

[11] Hryshko, D. (2012). "Labor income profiles are not heterogeneous: Evidence from income growth rates",*Quantitative Economics*, 3(2), 177-209.

[12] Hryshko, D. (2013): "*Excess smoothness of consumption in an Estimated Life-Cycle Model*", Manuscript

[13] Im, K. S.; M. H. Pesaran and Y. Shin. 2003. "Testing for Unit Roots in Heterogeneous Panels", *Journal of Econometrics*, 115(1), pp. 53-74

[14] Jappelli, T. and L. Pistaferri (2010), "The Consumption Response to Income Changes." *Annual Review of Economics*, 2(2), 479-506.

[15] Levin, A.; C. F. Lin and C. S. J. Chu. 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties", *Journal of Econometrics*, 108(1), pp. 1-24.

[16] Meghir, C. and L. Pistaferri (2004), "Income variance dynamics and heterogeneity." *Econometrica* 72(1): 1-32.

[17] Moffitt, R. A. and P. Gottschalk (2002), "Trends in the transitory variance of earnings in the United States." *Economic Journal* 112(478): C68-C73.

[18] Moffitt, R. A. and P. Gottschalk. 2012. "Trends in the Transitory Variance of Male Earnings, Methods and Evidence", *Journal of Human Resources*, 47(1), pp. 204-36.

[19] Morley, J., C.R. Nelson and E. Zivot (2003): "Why are the Beveridge-Nelson and Unobserved-Components decomposition so different?" *Review of Economic and Statistics*. Vol 85(2) pp. 235-243..

[20] Oh, K.H., E. Zivot and D. Creal (2008), "The relation between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics", *Journal of Econometrics*, 146, 207-219.

[21] Pistaferri, L. (2001), "Superior Information, Income Shock and Permanent Income Hypothesis", *Review of Economic and Statistics*, 83(3), pp. 465-476.

[22] Proietti, T (2006), "Temporal Disaggregation by state-space methods: Dynamic regression revisited", *Econometric Journal*, 9, 357-372.

A The ARMA(1,2) case with zero covariance.

In this appendix we consider the $ARMA(1,2)$ model. For simplicity we assume a process with mean zero and no trend and again we assume that

$0 < \rho \leq 1$. The $ARMA(1,2)$ is given by:

$$y_t = \rho y_{t-1} + \xi_t + \theta_1 \xi_{t-1} + \theta_2 \xi_{t-2}. \quad (31)$$

This model can under certain parameter restrictions be written as a PT model with uncorrelated shocks, where the transitory shock τ_t is an $MA(1)$ process with parameter λ:

$$\begin{aligned} \xi_t + \theta_1 \xi_{t-1} + \theta_2 \xi_{t-2} &= \varepsilon_t + \lambda \varepsilon_{t-1} - \rho(\varepsilon_{t-1} + \lambda \varepsilon_{t-2}) + \eta_t \\ &= \varepsilon_t + (\lambda - \rho)\varepsilon_{t-1} - \rho\lambda\varepsilon_{t-2} + \eta_t \end{aligned}$$

Taking covariances gives three equations that give the mapping between the two sets of parameters:

$$\begin{aligned} (1 + \theta_1^2 + \theta_2^2)\nu^2 &= (1 + (\lambda - \rho)^2 + (\rho\lambda)^2)\sigma_\varepsilon^2 + \sigma_\eta^2 \\ \theta_1(1 + \theta_2)\nu^2 &= (\lambda - \rho)(1 - \lambda\rho)\sigma_\varepsilon^2 \\ \theta_2 \nu^2 &= -\lambda\rho\sigma_\varepsilon^2 \end{aligned}$$

We consider the unit root case: $\rho = 1$ and we restrict θ_1, θ_2 and λ to lie within -1 and 1. The equations are given by:

$$\begin{aligned} (1 + \theta_1^2 + \theta_2^2)\nu^2 &= 2(1 + \lambda^2 - \lambda)\sigma_\varepsilon^2 + \sigma_\eta^2 \\ \theta_1(1 + \theta_2)\nu^2 &= -(\lambda - 1)^2\sigma_\varepsilon^2 \\ \theta_2 \nu^2 &= -\lambda\sigma_\varepsilon^2 \end{aligned}$$

From the middle equation we can immediately see that θ_1 has to be negative. θ_2 can be positive, zero and negative depending on the sign of λ. We now consider three cases

	$\lambda < 0$	$\lambda = 0$	$\lambda > 0$
θ_1	< 0	< 0	< 0
θ_2	> 0	$= 0$	< 0

Proposition 8 *The ARMA(1,2) model with $\rho = 1$ has an uncorrelated PT representation if the parameters θ_1 and θ_2 satisfy the following restrictions:*

$$\begin{aligned} -1 &\leq \theta_1 \leq \frac{-4\theta_2}{(1+\theta_2)}, \ \theta_2 > 0 \\ \theta_1 &\leq 0, \theta_2 \leq 0. \end{aligned}$$

The PT parameters are given by if $\theta_2 \neq 0$

$$\begin{aligned} \sigma_\eta^2 &= (1+\theta_1+\theta_2)^2 \nu \\ \lambda &= \frac{1}{2\theta_2}\left(\theta_1 + 2\theta_2 + \theta_1\theta_2 \pm \sqrt{\theta_1(\theta_2+1)(\theta_1+4\theta_2+\theta_1\theta_2)}\right) \\ \sigma_\varepsilon^2 &= \frac{\theta_2 \nu}{-\lambda} = -\frac{2\theta_2^2}{\left(\theta_1 + 2\theta_2 + \theta_1\theta_2 \pm \sqrt{\theta_1(\theta_2+1)(\theta_1+4\theta_2+\theta_1\theta_2)}\right)} \end{aligned}$$

If $\theta_2 = 0$ then $\lambda = 0$ and $\sigma_\varepsilon^2 = -\theta_1\nu^2$.

In figure 4, we show the parameter space of the ARMA(1,2) with $\rho = 1$ that is consistent with a PT-representation. On the figure we also show the ARMA(1,2) parameter implied by the Meghir and Pistaferri (2004) estimation.

Corollary 9 *In the unit root model $\rho = 1$ there is* no transitory shocks *if $\theta_1 = \theta_2 = 0$. If $\theta_1 + \theta_2 = -1$ there is* no permanent shocks.

Moving on to the more general case with $\rho \in (0,1)$ and θ_1, θ_2 and $\lambda \in (-1,1)$ We now consider the different cases

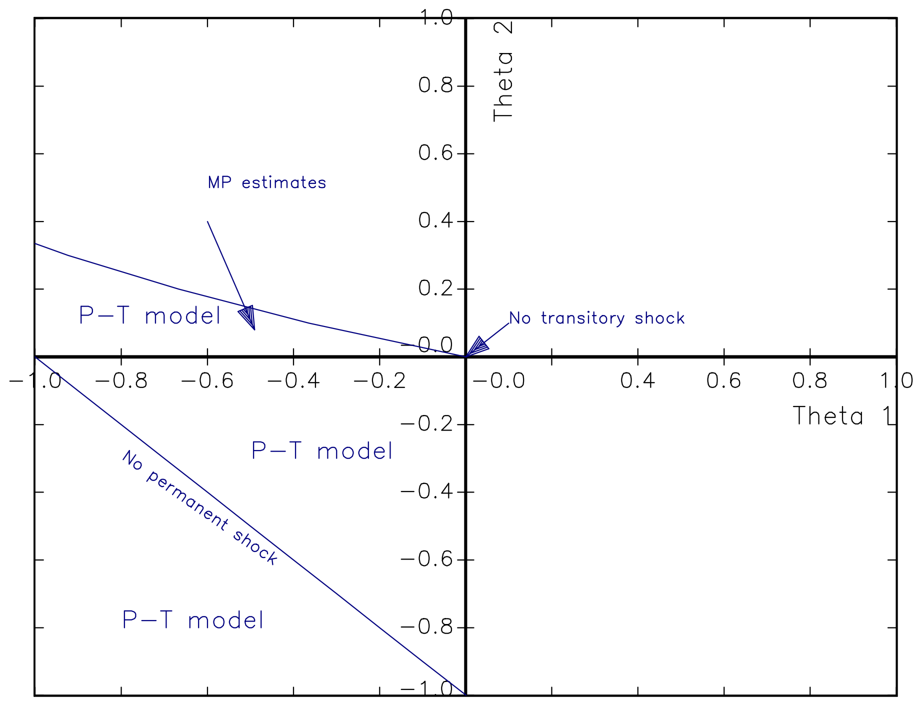


Figure 4: The restrictions of the ARMA(1,2) parameters when $\rho = 1$

	$\lambda < \rho$	$\lambda = \rho$	$\lambda > \rho$
$\lambda < 0$	$\theta_1 < 0$ $\theta_2 > 0$		
$\lambda = 0$	$\theta_1 < 0$ $\theta_2 = 0$		
$\lambda > 0$	$\theta_1 < 0$ $\theta_2 < 0$	$\theta_1 = 0$ $\theta_2 < 0$	$\theta_1 > 0$ $\theta_2 < 0$

From the table above one can conclude that if both θ_1 and θ_2 are positive then the earnings process cannot be represented with a standard (uncorrelated) PT model.

B The Beveridge-Nelson approach

We also use the Beveridge-Nelson approach to decompose the $ARMA(1,2)$ into a persistent shock and transitory shock.[18] However, it turns out that this decomposition is different than the PT model with uncorrelated shock. For simplicity we assume a process with mean zero and no trend and again we assume that $0 < \rho \leq 1$. The $ARMA(1,2)$ is given by

$$y_t = \rho y_{t-1} + \xi_t + \theta_1 \xi_{t-1} + \theta_2 \xi_{h-2}. \tag{32}$$

By using the time series representation we can write the $ARMA$ model as

$$A(L) y_t = \theta(L) \xi_t,$$

[18] The decomposition is strictly speaking not a BN decomposition but uses the same ideas.

where

$$A(L) = 1 - \rho L \text{ and } \theta(L) = 1 + \theta_1 L + \theta_2 L^2.$$

We use the BN decomposition of a general $ARMA$ model into a persistent and transitory component. Under the condition that the $A(L)$ and $\theta(L)$ have no common roots we can write $y_t = (A(L))^{-1}\theta(L)\xi_t$. Since the root in the $A(L)$ is $r = 1/\rho$, the condition of no common roots implies that $1/\rho$ can not be a root in $\theta(L)$:

$$\theta(1/\rho) = 1 + \theta_1/\rho + \theta_2/\rho^2 \neq 0,$$

We then make the decomposition as

$$y_t = (A(L))^{-1}\theta(L)\xi_t = \sigma(A(L))^{-1}\xi_t + (A(L))^{-1}(\theta(L) - \sigma)\xi_t,$$

where σ is a scalar which indicate how much of the shock which can be attributed to a persistent shock. The first term $\sigma(A(L))^{-1}\xi_t$ is the persistent part of the process:

$$\sigma(A(L))^{-1}\xi_t = \sigma \sum_{i=0}^{\infty} \rho^i \xi_{t-i}$$

In order to make the second term a transitory shock we require that this process is an $MA(q)$ process. To be this we need that $A(L))$ and $(\theta(L) - \sigma)$ have common roots; i.e. that $1/\rho$. should be a root in $(\theta(L) - \sigma)$

$$\begin{aligned} 1 - \sigma + \theta_1/\rho + \theta_2/\rho^2 &= 0 \\ 1 + \theta_1/\rho + \theta_2/\rho^2 &= \sigma \end{aligned}$$

Using the factorization of $(\theta(L) - \sigma)$ we get

$$(\theta(L) - \sigma) = (L - 1/\rho)(\theta_1 + \theta_2(L + 1/\rho)).$$

The transitory shock is given by

$$\begin{aligned}(A(L))^{-1}(\theta(L)-\sigma)\xi_t &= (1-\rho L)^{-1}(L-1/\rho)(\theta_1+\theta_2(L+1/\rho))\xi_t \\ &= -1/\rho(\theta_1+\theta_2(L+1/\rho))\xi_t \\ &= -1/\rho*(\theta_1+\theta_2/\rho)\xi_t - 1/\rho*\theta_2\xi_{t-1}.\end{aligned}$$

The persistent shock is given by

$$\eta_t = (1+\theta_1/\rho+\theta_2/\rho^2)\xi_t$$

The decomposition is given by

$$y_t = (1+\theta_1/\rho+\theta_2/\rho^2)\sum_{i=0}^{\infty}\rho^i\xi_{t-i} - 1/\rho*(\theta_1+\theta_2/\rho)\xi_t - 1/\rho*\theta_2\xi_{t-1}$$

If we restrict the scalar to $0 \leq \sigma \leq 1$, which implies that σ can be interpreted as the part of the shock which is the persistent part we can derive the restrictions on the $ARMA$ parameters:

$$\begin{aligned}0 &< 1+\theta_1/\rho+\theta_2/\rho^2 < 1. \qquad (33)\\ -1 &< \theta_1/\rho+\theta_2/\rho^2 < 0\end{aligned}$$

If $1+\theta_1/\rho+\theta_2/\rho^2 = 0$ we have no persistent shock and if $1+\theta_1/\rho+\theta_2/\rho^2 = 1$ we have no transitory shocks.

The main difference between the time series approach and the PT model is that the time series approach does not require that the persistent and transitory shocks are uncorrelated. In the time series approach the covariance of the shocks is given by this

$$Cov(\eta_t, 1/\rho*(\theta_1+\theta_2/\rho)\xi_t - 1/\rho*\theta_2\xi_{t-1}) = -1/\rho*(\theta_1+\theta_2/\rho)*(1+\theta_1/\rho+\theta_2/\rho^2)*\nu^2.$$

Given the restrictions (33) the covariance will always be positive but otherwise the the covariance will be negative.

C Decomposition of shocks

We start by simulating a PT model with uncorrelated shocks. The parameters of this model is given by $(\rho, \sigma_\eta^2, \sigma_\tau^2)$. We simulate a sequence of independent transitory and persistent shocks: $(\tau_1^0, ..., \tau_T^0)$ and $(\eta_1^0, ..., \eta_T^0)$ and the process is given by

$$\begin{aligned} y_t &= p_t + \tau_t^0, \\ p_t &= \rho p_{t-1} + \eta_t^0 \ t > 1 \\ p_1 &= \eta_1^0, \tau_1^0 = 0 \end{aligned}$$

We know that we can write this process as

$$\begin{aligned} y_t - \rho y_{t-1} &= \xi_t + \theta \xi_{t-1} \quad t > 1 \\ y_1 &= \xi_1 \end{aligned}$$

where the parameters in the model are ρ, θ, ν^2 can be found as

$$\begin{aligned} \theta &= \frac{1}{2\rho\sigma_\tau^2}\left[-(1+\rho^2)\sigma_\tau^2 - \sigma_\eta^2 + \sqrt{(1-\rho^2)^2\sigma_\tau^4 + \sigma_\eta^4 + 2(1+\rho^2)\sigma_\tau^2\sigma_\eta^2}\right] \\ \nu^2 &= (-\rho/\theta) * \sigma_\tau^2 \end{aligned}$$

We can then recursively determine the shocks in the $ARMA(1,1)$ model

$$\begin{aligned} \xi_1 &= y_1 \\ \xi_t &= y_t - \rho y_{t-1} - \theta \xi_{t-1} \quad t > 1 \end{aligned}$$

We can then define the transitory and persistent shock by using the BN decomposition:

$$\begin{aligned}
\eta_1^P &= y_1, \tau_1^P = 0 \\
\tau_t^P &= \frac{-\theta}{\rho}\xi_t \quad t > 1 \\
\eta_t^P &= (1 + \frac{\theta}{\rho})\xi_t \quad t > 1
\end{aligned}$$

Figure 3 is generated for $\rho = 1, \sigma_\tau^2 = \sigma_\eta^2 = 1$.

D Empirical details.

D.1 The auxiliary parameters

We have five heterogeneous model parameters $(\mu, \alpha, \rho, \theta, \nu)$. We wish to define IRB ap's that are 'bound' to these.

The original data is denoted y_{it} for $i = 1, ..H$ and $t = t_{1i}, ..t_{Ti}$ where the latter values are the first and last periods for i.

1. Take deviations about the mean: $x_{it} = y_{it} - \bar{y}_i.$.

2. De-trend by regressing the deviations about the mean on a trend:

$$x_{it} = b_{1i} * t + u_{1it} \tag{34}$$

   and record the estimated residuals $\hat{u}_{1it} = x_{it} - \hat{b}_{1i}t$ for $t = t_{1i}, ..t_{Ti}$.

3. Regress these residuals on their lagged values:

$$\hat{u}_{1it} = b_{2i}\hat{u}_{1i,t-1} + u_{2it} \tag{35}$$

and record the estimated residuals $\hat{u}_{2it} = \hat{u}_{1it} - \hat{b}_{2i}\hat{u}_{1i,t-1}$ for $t = t_{1i} + 1, ..t_{Ti}$. The estimate b_{2i} is bound the auto-regressive parameter.

4. Perform the following regression:

$$x_{it} - b_{2i}x_{it-1} = b_{3i} * t + u_{3it} \tag{36}$$

The estimated $\hat{b}_{3i}$'s are bound to the trend parameter.

5. Calculate:

$$\hat{b}_{4i} = \left(1 - \hat{b}_{2i}\right)\bar{y}_i - \hat{b}_{3i} * t \tag{37}$$

The parameters are bound to the intercept of the process.

6. Calculate residuals for $t = t = t_{1i} + 1, ..t_{Ti}$:

$$\hat{u}_{3it} = y_{it} - \hat{b}_{4i} - b_{2i} * y_{i,t-1} - \hat{b}_{3i} * t \tag{38}$$

and record the standard deviation and auto-correlation (denoted $\hat{b}_{5i}$ and $\hat{b}_{6i}$, respectively). These are for the variance and MA parameters.

The ap's for the five model parameters $(\mu, \alpha, \rho, \theta, \nu)$ are, respectively, $\left(\hat{b}_{4i}, \hat{b}_{3i}, \hat{b}_{2i}, \hat{b}_{6i}, \hat{b}_{5i}\right)$. We also record the initial value $y_{i1}$and take means, standard deviations and correlations of the six values for each worker.

D.2 Parameter estimates

	estimate	std error	t-value
ϕ_1	-0.089	0.314	0.283
ϕ_2	-1.844	0.033	55.668
ϕ_3	-0.033	0.063	0.517
ϕ_4	1.518	0.402	3.772
ϕ_5	-0.448	0.110	4.088
c_0	-3.581	0.997	3.590
c_1	0.010	6.372	0.002
ψ_{11}	-1.776	1.084	1.639
ψ_{21}	0.488	0.105	4.649
ψ_{22}	-1.620	1.217	1.331
ψ_{31}	-0.051	0.029	1.771
ψ_{32}	-0.017	0.032	0.523
ψ_{41}	-1.096	0.504	2.176
ψ_{42}	-0.610	0.710	0.860
ψ_{51}	0.367	0.179	2.050
ψ_{52}	-0.303	0.212	1.428
a_0	-0.026	0.442	0.059
a_1	0.313	0.850	0.369